TRANSACTIONS
OF THE
AMERICAN MATHEMATICAL SOCIETY

EDITED BY

EINAR HILLE C. C. MAC DUFFEE OSCAR ZARISKI

VOLUME 49
JANUARY TO JUNE, 1941

PUBLISHED BY THE SOCIETY
MENASHA, WIS., AND NEW YORK
1941

Chester C. Corbin Library Fund
Bound June 1941

Composed, Printed and Bound by
The Collegiate Press
George Banta Publishing Company
Menasha, Wisconsin

TABLE OF CONTENTS

VOLUME 49, JANUARY TO JUNE, 1941

TWO-TO-ONE TRANSFORMATIONS ON 2-MANIFOLDS

BY
VENABLE MARTIN AND J. H. ROBERTS

Introduction. An *exactly* 2-*to*-1 transformation is one for which every inverse image set consists of exactly 2 points. This notion was introduced by O. G. Harrold(1), who showed that no such continuous transformation could be defined over an arc. This result has been extended(2) to the case of the closed 2-cell. Further results concerning these transformations have been obtained by Harrold and by P. W. Gilbert(3). The present paper is concerned with continuous 2-to-1 transformations defined on a compact 2-manifold, with or without bounding curves. The problem of the existence of such a transformation is solved, and the collection of all image spaces is determined. A precise statement of the main results is given below.

Throughout this paper the letter M will be used to denote a compact 2-manifold (absolute), or else a compact 2-manifold with boundary, the boundary consisting of a finite number of mutually exclusive simple closed curves. The set M will be considered as the whole space. T will denote some exactly 2-to-1 continuous transformation defined over M. The set of inverse images under T is(4) an upper semi-continuous collection G filling M, and every element of G is a pair of points. For each $x \varepsilon M$ let $s(x)$ be the point such that the pair $x, s(x)$ is an element of the collection G. Let $f(x) = \rho(x, s(x))$, where ρ is the metric on M. Let K denote the set of all points $x \varepsilon M$ at which f is continuous, and let L denote the subset of K consisting of those points x such that f is continuous both at x and at $s(x)$. If x is a point, then x' will denote $s(x)$; and if C is any point set, then C' will denote $s(C)$. A point set C will be called *integral* if $C' = C$.

The term *n-cell* ($n = 0, 1, 2$) will denote a closed n-cell except where the context indicates the contrary. If β denotes a closed n-cell, then β^0 will denote the open n-cell whose closure is β. If A denotes a complex, then A^* denotes

Presented to the Society, February 24, 1940; received by the editors March 25, 1940.

(1) *The non-existence of a certain type of continuous transformation*, Duke Mathematical Journal, vol. 5 (1939), pp. 789–793. See also abstracts by Harrold in the Bulletin of the American Mathematical Society, vol. 46 (1940), pp. 43, 44.

(2) J. H. Roberts, *Two-to-one transformations*, Duke Mathematical Journal, vol. 6 (1940), pp. 256–262. This paper will be referred to hereafter as *Transformations*.

(3) See abstracts in the Bulletin of the American Mathematical Society, vol. 45 (1939), p. 835, and vol. 46 (1940), pp. 42, 43.

(4) This follows readily from the compactness of M. However, the corresponding statement in *Transformations* does not follow from the continuity of T, and must be taken as an extra hypothesis in those theorems of that paper where M was not assumed to be compact. This does not affect the main result of that paper.

the point set covered by the complex A. If A is a complex, then $\chi(A)$ will denote $-\alpha_0+\alpha_1-\alpha_2$, where α_i is the number of i-cells in A. This is the negative of the Euler characteristic as given by Alexandroff-Hopf(6). We assume a metric ρ on M having the following property: If a and b are points of M, then for any $\epsilon>0$ there is an arc ab whose diameter is less than $\rho(a, b)+\epsilon$. If σ is any metric on M, then a metric ρ with the above property can be obtained by taking $\rho(a, b)$ equal to the g.l.b. of the diameters of arcs joining a and b in M.

The principal results of the paper are as follows:

A necessary and sufficient condition that there be a T defined over M is that $\chi(M)$ be even. If $T(M)=B$, then $\chi(M)=2\chi(B)$. Let B_k denote a space which can be obtained from a compact manifold with k bounding curves ($k=0, 1, 2, \cdots$), by the identification by pairs of a finite number of interior points of the manifold. A compact manifold M with n bounding curves ($n=0, 1, 2, \cdots$), and a space B_k are said to be *properly related* if (1) $\chi(M)=2\chi(B_k)$ and (2) $\frac{1}{2}n\leqq k\leqq n$. *Given a manifold M and a space B_k, a necessary and sufficient condition that there be a T defined over M such that $T(M)=B_k$ is that M and B_k be properly related.*

These results are obtained in Part II. In order to obtain these results, it is essential to determine the nature of the discontinuities of f and the topological character of the set $M-K$. This is done in Part I.

PART I

LEMMA 1. *If the point p is not on the boundary of M (i.e., if p has a 2-cell neighborhood), then p is in L if it is in K.*

Proof. If p is in K, then a sufficiently small open 2-cell containing p is mapped topologically by s into an open 2-cell containing $s(p)$. Since an open 2-cell in M is necessarily open in M, and since s has period 2, it follows that $s(p)$ is also in K, whence p is in L.

THEOREM 1. *The set L is dense and open in M.*

Proof. Since K is dense and open (see *Transformations*), it follows from Lemma 1 that L is dense. If p is in L, then p and $s(p)$ are in the open set K. In view of the upper semi-continuity of the collection G it follows that if x is sufficiently close to p or to $s(p)$, then x and $s(x)$ are in K, and therefore in L. Thus L is open.

LEMMA 2. *If a simple closed curve J bounds an open 2-cell U, and R is any region which contains all of J except possibly one point, then $R\cdot U$ is connected.*

Proof. Let x and y be any two points of $R\cdot U$ and let xy be an arc joining x and y in R. If $xy\cdot J=0$, then the arc xy lies in $R\cdot U$. If $xy\cdot J\neq 0$, then let z

(6) *Topologie* I, Berlin, 1935, p. 214.

and w be points on J and on the arc xy in the order $xzwy$ (possibly $z=w$) such that no point of J precedes z or follows w on the arc $xzwy$. In view of the hypothesis on J there is an arc zvw belonging to $J\cdot R$. Since finally R is open and contains zvw, there are points z_1 and w_1 on the arc xy in the order xz_1zww_1y and an arc z_1w_1 lying in $R\cdot U$. Then the arc $xz_1+z_1w_1+w_1y$ joins x and y and lies in $R\cdot U$. Since each pair x, y of $R\cdot U$ lies on an arc in $R\cdot U$, this set is connected.

LEMMA 3. *If H is a closed 2-cell in M, then H does not contain 5 arcs ad_ia' $(i=1, \cdots, 5)$ such that* (1) *each 2 of these arcs have only their end-points a and a' in common,* (2) *$\sum ad_ia'$ is in L except for the points d_3, d_4, and d_5, which are not in K,* (3) *$d_2=d_1'$ and $s(ad_1a')=a'd_2a$, and* (4) *for $i=3$, 4, and 5, $s(ad_i-d_i)=a'd_i-d_i$.*

Proof. Suppose the lemma is false and there exist five arcs ad_ia' in a closed 2-cell H and properties (1), (2), (3) and (4) of the lemma hold. Now for $i=3$, 4, or 5, as $x\rightarrow d_i$ on the arc ad_i the point $s(x)\rightarrow d_i$ on the arc $a'd_i$. Let J denote the simple closed curve $ad_1a'd_2a$, and let U denote its interior with respect to H. Let R be the component of L containing J. We consider three cases.

Case 1. *At least two of the points d_3, d_4, d_5 are in U.* Suppose that d_3 and d_4 are in U. Then the arc ad_3a' lies in $U+a+a'$, and $U=U_1+U_2+ad_3a'-a-a'$, where U_1 and U_2 are open 2-cells bounded by $ad_1a'd_3a$ and $ad_2a'd_3a$, respectively. We suppose, without loss of generality, that d_4 is in U_2. Now let R_0, R_1, and R_2 denote respectively $R\cdot U$, $R\cdot U_1$, and $R\cdot U_2$. By Lemma 2 the sets R_0, R_1, and R_2 are connected. Furthermore, they are open subsets of the open 2-cell U. Hence $s(R_0)$, $s(R_1)$, and $s(R_2)$ are connected open sets, since s is topological over R, and an open 2-cell in M is necessarily open in M. Since $s(J)=J$, and $s(ad_ia'-d_i)=a'd_ia-d_i$ $(i=3, 4, 5)$, it follows that $s(R_0)$, $s(R_1)$, and $s(R_2)$, respectively do not intersect the boundary of U, U_1 and U_2. That is, either $s(R_0)\subset U$ or else $s(R_0)$ is in $M-\overline{U}$. But the second possibility is ruled out since $s(ad_4a'-d_4)=ad_4a'-d_4$, and this set is in R_0. Therefore $s(R_0)\subset U$. Likewise $s(R_2)\subset U_2$, for d_4 is in U_2. But now let $x_1, x_2, \cdots$ be points in R_2 such that $x_n\rightarrow d_2$. Since d_2 is in L and $s(d_2)=d_1$, $s(x_n)\rightarrow d_1$. But $s(x_n)\in U_2$ and d_1 is not in $\overline{U}_2$. This is a contradiction.

Case 2. *Exactly one of the points d_3, d_4, d_5 is in U.* Suppose that $d_5\in U$ and d_3 and d_4 are not in U. The sum of ad_3a' and one of the two arcs ad_1a', ad_2a' is a simple closed curve bounding a 2-cell U_1 which lies in H and contains the other of these arcs, except for end-points. We suppose that the boundary of U_1 is $ad_3a'd_2a$, and let U_2 be the 2-cell in H bounded by $ad_3a'd_1a$. To summarize, $H\supset\overline{U}_1$, and $U_1=U+U_2+ad_1a'-a-a'$. Let R_0, R_1, and R_2 denote respectively $R\cdot U$, $R\cdot U_1$, and $R\cdot U_2$. Then it follows that $s(R_0)\subset U$ and $s(R_1)\subset U_1$, since both U and U_1 contain the set $ad_5a'-a-a'$, and $s(ad_5a'-d_5)=ad_5a'-d_5$. Now let $x_1, x_2, \cdots$ be points in R_2 converging to d_1.

Then $s(x_1), s(x_2), \cdots$ are points in U_1 (for $R_2 \subset R_1$) converging to d_2. But for n sufficiently large $s(x_n)$ is in U, hence in R_0. But $s(R_0) \subset U$, and $s(s(x_n)) = x_n$. Thus x_n is in both U and U_2 and we have a contradiction.

Case 3. *None of the points d_3, d_4, d_5 is in U.* Then there is a 2-cell in H bounded by the sum of one arc from the set $\{ad_ia'\}$ $(i=1, 2)$ and one arc from the set $\{ad_ia'\}$ $(i=3, 4, 5)$, and containing one of d_1, d_2, and one of d_3, d_4, d_5. For definiteness we may suppose there are 2-cells U, U_1, and U_2 bounded respectively by curves $ad_1a'd_2a$, $ad_3a'd_2a$, and $ad_3a'd_1a$, such that d_4 is in U_2 and $U_1 = U + U_2 + ad_1a' - a - a'$. Let $R_2 = R \cdot U_2$. Since d_4 is in U_2, it follows by earlier arguments that $s(R_2) \subset U_2$. Let $x_1, x_2, \cdots$ be points in R_2 converging to d_1. Then $s(x_n) \to d_2$. But d_2 is not in $\overline{U}_2$ while $s(x_n) \subset U_2$. This contradiction completes the proof of the lemma

LEMMA 4. *Suppose q is a point, H is a closed 2-cell, V is an open set, and ϵ is a positive number, and the following properties hold:*

(1) *H contains a neighborhood of q;*

(2) *$V \supset q$;*

(3) *q is a limit point of $M-K$, and*

(4) *if $p \in K \cdot V$, $f(p) < \epsilon$, and pt is any arc in V and in $K+t$, where t is not in K, then $f(x) \to 0$ as $x \to t$ on the arc pt.*

Then it follows that there exists an $\epsilon_1 > 0$ such that no arc cc' lies in $K \cdot S(q, \epsilon_1)$.

Proof. Suppose the lemma is false. Then for each positive integer n there is an arc c_nc_n' which is a subset of each of the sets H, K, V, and $S(q, 1/n)$, and for n sufficiently large the arc $s(c_nc_n')$ is in $H \cdot K \cdot V$. The set $c_nc_n' + s(c_nc_n')$ contains[6] a simple closed curve J, the sum of two arcs utu' and $u't'u$ such that $s(utu') = u't'u$. Then J bounds an open 2-cell H_1 which is a subset of H. Let R be the component of K which contains J. Then by Lemma 2 $R \cdot H_1$ is connected. We want to get a simple closed curve $ad_1a'd_2a$ lying in $R \cdot H_1$ and such that $s(ad_1a') = a'd_2a$. This will follow readily if we prove that $s(R \cdot H_1) = R \cdot H_1$, for then any point p in this set can be joined to p' by an arc pp' in this set, and some subset of $pp' + s(pp')$ will be the desired curve. If there is no point of $M-K$ in H_1, then s is topological over $\overline{H}_1$, and $s(\overline{H}_1)$ is a closed 2-cell with J for boundary. Either $s(H_1) \subset H_1$ or $s(H_1) \subset M - \overline{H}_1$. But the first possibility is ruled out, because under it s is a topological mapping of the closed 2-cell $\overline{H}_1$ into itself which has no fixed point. Under the second possibility $\overline{H}_1 + s(\overline{H}_1)$ is a sphere, hence is M. But $\overline{H}_1 + s(\overline{H}_1)$ is in K, contrary to the fact that there are points in $M-K$ (e.g., the point q). Thus there is some point of $M-K$ in H_1. Join a point p of $R \cdot H_1$ to a point t of $M-K$ in H_1 by an arc pt lying in H_1. Then $f(x) \to 0$ as $x \to t$ on the arc pt, and therefore for x near enough to t, $s(x)$ is also in H_1. Hence $s(R \cdot H_1) \subset H_1$.

Now by Lemma 1, $s(R \cdot H_1) \subset K$. It follows that $s(R \cdot H_1) = R \cdot H_1$, and the desired simple closed curve exists. That is, there is a simple closed curve

(6) See *Transformations*, §8, for a proof.

$ad_1a'd_2a$ lying in H_1 and in K, and such that $s(ad_1a')=a'd_2a$. By Lemma 1, this curve lies in L. There is an arc joining a to some point of $M-K$, such that this arc is in $H\cdot V$, has only the point a on the simple closed curve $ad_1a'd_2a$, and has no point, except possibly an end-point, on the boundary of M. On this arc let d_3 be the first point of $M-L$. Then d_3 is in $M-K$. For if d_3 is the other end-point it is by definition in $M-K$. In the other case d_3 is not on a bounding curve of M and hence by Lemma 1 is in $M-K$ if it is in $M-L$. Then $f(x)\to 0$ as $x\to d_3$ on the arc ad_3. It follows from Theorem 3 of *Transformations*(7) that $ad_3+s(ad_3-d_3)$ contains two arcs, atd_3 and $a't'd_3$ which have only d_3 in common and such that $s(atd_3-d_3)=a't'd_3-d_3$. Denote the sum $atd_3+d_3t'a'$ by ad_3a'.

In a similar way we obtain successively arcs ad_4a' and ad_5a' such that for $i=4, 5$, d_i is in $M-K$ but $ad_ia'-d_i$ is in L, ad_ia' has only a and a' in common with the sum of the other four arcs ad_ja', and $s(ad_i-d_i)=a'd_i-d_i$. But then the five arcs ad_ia' $(i=1, \cdots, 5)$ have the properties stated in Lemma 3, and we have reached a contradiction.

LEMMA 5. *Suppose H is a closed 2-cell which contains a neighborhood V of a point q, pq is an arc in V and in $L+q$ such that $f(x)\to 0$ as $x\to q$ on the arc pq. Let R be the component of $L\cdot V$ which contains p, and suppose $R\cdot s(R)\cdot V=0$. Let ϵ be any positive number. Then there exists in V an open set W with boundary J such that*

(1) *$W\supset q$ and $W+J$ is of diameter less than ϵ, and $\overline{W}$ is a closed 2-cell;*

(2) *if q is on a bounding curve of M, then J is an arc xax', where x and x' are on a bounding curve of M, a is in $M-K$, $xa-a$ is in R, and $s(xa-a)=x'a-a$;*

(3) *if q is not on a bounding curve of M, then J is a simple closed curve $axbx'a$, where a and b are in $M-K$, $axb-a-b$ is in R, and $s(axb-a-b)=ax'b-a-b$.*

Proof. The proof given for Theorem 5 of *Transformations* requires only a trivial change in order to apply here.

LEMMA 6. *Suppose q_1 is a point, H is a closed 2-cell, V is an open set and cq_1 is an arc, and the following properties hold:*

(1) *$H\supset V\supset cq_1$;*

(2) *if $x\in H$, then either $f(x)<\epsilon$ or $f(x)>3\epsilon$, where $4\epsilon=f(q_1)$;*

(3) *$c\in L$ and $f(c)<\epsilon$;*

(4) *there is no arc connecting any point d to d' and lying in $V\cdot K$; and*

(5) *if ef is any arc lying in $V\cdot L+f$, where $f(e)<\epsilon$ and f is not in K, then $f(x)\to 0$ as $x\to f$ on the arc ef.*

Then there exists an arc from c to q_1 and lying in $V\cdot L+q_1$.

---

(7) Theorem 3 of *Transformations* is false as stated. The proof given is based on the assumption that K is an integral set, i.e., that $s(K)=K$. Now L is an integral set, and the argument given suffices to prove the theorem as stated if K is replaced by L. In our application the the arc ad_3 lies in $L+d_3$, hence the modified theorem applies.

Proof. Let R_1 be the component of $L\cdot V$ that contains c. Then $R_1\cdot s(R_1)=0$. Let E denote the set of all x in $R_1+s(R_1)$ for which $f(x)<\epsilon$. Let t be the last point of $\overline{E}$ on the arc cq_1.

Suppose that t is not accessible from R_1 by an arc ct such that $f(x)\to 0$ as $x\to t$ on ct. Then t is not accessible by any arc ut lying in $R_1+s(R_1)+t$ and containing a point of E. Then we will show that there is an infinite sequence $c_1, c_2, c_3, \cdots$ such that (1) $c_n \varepsilon R_1$ and $f(c_n)<\epsilon$, (2) $c_n\to t$ as $n\to\infty$, and (3) there is a fixed positive δ such that every arc joining c_i and c_j $(i\neq j)$ in R_1 has diameter greater than δ. To prove this assertion consider the following hypothesis:

Given any positive number β and any component R_β of $R_1\cdot S(t, \beta)$ having t on its boundary, it is true that for every k there is some component of $R_\beta\cdot S(t, 1/k)$ which has t on its boundary. If this is true, then it follows that t is accessible from R_1 by an arc ct such that $f(x)\to 0$ as $x\to t$ along ct. But this contradicts a supposition made above. Hence the above hypothesis is false. This means that there is a positive number β such that there is an infinite set $R_1^1, R_1^2, R_1^3, \cdots$ of components of $R_1\cdot S(t, \beta)$ such that, for every i, t is not a limit point of R_1^i, but t is a limit point of $\sum_{i=1}^{\infty}R_1^i$. And this implies that the sequence $c_1, c_2, c_3, \cdots$ exists.

There exist three open sets W_1, W_2, and W_3 containing t and bounded respectively by P_1, P_2, and P_3, these being simple closed curves or arcs([8]), and there is an integer N, such that

(1) $V\supset \overline{W}_1$ and $W_i\supset \overline{W}_{i+1}$ $(i=1, 2)$;

(2) if $n>N$, there is an arc $c_nd_ne_n$ in R_1 and in $\overline{W}_1$, where e_n and d_n are on the boundaries of W_1 and W_2, respectively, and c_n is in W_3; and

(3) if $n>N$, $m>N$, and $n\neq m$, then no component of $R_1\cdot\overline{W}_1$ contains both c_n and c_m.

Suppose $n>N$. Let x_n and y_n be the first points of the boundary of R_1 on the circle P_2 starting from d_n in the two senses, and let $x_nd_ny_n$ denote the indicated arc of the circle P_2. Then $s(x_nd_ny_n-x_n-y_n)+x_n+y_n$ is an arc $x_nd_n'y_n$ in $s(R_1)+x_n+y_n$. Since $x_nd_n'y_n$ does not intersect $c_nd_ne_n$ (because $R_1\cdot s(R_1)=0$), there is a positive γ independent of n such that $d(x_nd_n'y_n)>\gamma$. But $d(x_nd_ny_n)\to 0$ as $n\to\infty$. If we drop to a subsequence, we may suppose $\limsup_{n\to\infty} x_nd_ny_n$ is a point r. Then $\limsup_{n\to\infty} x_nd_n'y_n$ is contained in $r+s(r)$. But this is clearly impossible, and we have thus proved that t is accessible by an arc ct lying in R_1+t, and therefore in $V\cdot L+t$. It follows that $f(x)\to 0$ as $x\to t$ on the arc ct.

Suppose now that $t\neq q_1$. Choose $\epsilon_1<\rho(t, q_1)$. Let W and J be sets given by Lemma 5, where t and the arc ct replace q and the arc pq, and ϵ_1 replaces ϵ. Then the arc tq must contain a point r $(r\neq t)$ on J. But J lies in $\overline{E}$, and t is the last point of $\overline{E}$ on the arc cq_1. This contradiction proves Lemma 6.

([8]) If t is on a bounding curve of M, then P_i is an arc with end-points on this bounding curve.

THEOREM 2. *If q is a point in $M-K$ and pq is an arc in $K+q$, then $f(x)\to 0$ as $x\to q$ on the arc pq.*

Proof. We suppose the theorem is false. Then there is an arc pq lying in $K+q$, q not in K, such that $f(x)\to f(q)$ as $x\to q$ on the arc pq. Then by Theorem 4 of *Transformations* there is an open set U, an arc p_1q_1, and a positive ϵ, such that

(1) $U\supset q_1$;

(2) $p_1q_1-q_1$ is in K, but q_1 is not in K;

(3) $f(x)\to f(q_1)=4\epsilon$ as $x\to q_1$ on the arc p_1q_1;

(4) if $x\in U$, then either $f(x)<\epsilon$ or $f(x)>3\epsilon$;

(5) if p_2q_2 is any arc in $U\cdot(K+q_2)$, and q_2 is not in K, and if $f(p_2)<\epsilon$, then $f(x)\to 0$ as $x\to q_2$ on the arc p_2q_2.

Let H be a closed 2-cell such that $U\supset H\supset q_1$, and H contains a neighborhood of q_1. By Lemma 4 there exists an $\epsilon_1>0$ such that if c is a point and $\rho(c, q_1)<\epsilon_1$, then there does not exist an arc of diameter less than ϵ_1 lying in K and joining c to $s(c)$. Let V denote an open set containing q_1 and lying in $H\cdot S(q_1, \epsilon_1)$. Since f is not continuous at q_1, there is a point c in $L\cdot V$ such that $f(c)<\epsilon$. There is an arc cq_1 in V. Then all the hypotheses of Lemma 6 are satisfied. Hence there is an arc cq_1 lying in $V\cdot L+q_1$. Then on this arc $f(x)\to 0$ as $x\to q_1$ (since $f(c)<\epsilon$).

Let ϵ_2 be the smaller of $\rho(p_1, q_1)$ and $\rho(c, q_1)$. Let H, V, q_1, cq_1, and ϵ_2, respectively, play the roles of H, V, p_1, cp_1, and ϵ in Lemma 5, and let J be the corresponding arc or simple closed curve having properties (2) and (3) of Lemma 5. Then J separates p_1 and q_1 and also c and q_1. Then for $x\in J$, $x\in K$ we have $f(x)<\epsilon$. But for x on p_1q_1 we have $f(x)>3\epsilon$. This is a contradiction, and the theorem is proved.

THEOREM 3. *If q is a limit point of $M-K$, then there is a positive number ϵ_1 such that there does not exist, for any point c, an arc joining c to $s(c)$ lying in K and in $S(q, \epsilon_1)$.*

Proof. Let H denote a closed 2-cell in M which contains a neighborhood of q. Let V be any open set containing q and let ϵ be any positive number. Then with the help of Theorem 2 it follows immediately that q, H, V, and ϵ have the properties stated in the hypothesis of Lemma 4. The number ϵ_1 given in the conclusion of Lemma 4 has the required property.

THEOREM 4. *If R is a component of K and q is on the boundary of R, then q is arc-wise accessible from R.*

Proof. If q is not a limit point of $M-K$, the result is obvious. If q is a limit point of $M-K$, then it is possible, with the help of Theorems 2 and 3, to define H, V, cq, having properties as stated in the hypothesis of Lemma 6, with the additional hypothesis that $c\in R$. Then the arc cq given by Lemma 6 will lie in $R+q$.

THEOREM 5. *The set L is identical with K.*

Proof. It is sufficient to show that if p is in K, then $s(p)$ is in K. This result was shown in Lemma 1 except for the case where p is on a bounding curve of M. Suppose then that p is in K and on a bounding curve of M. There is a simple closed curve J bounding an open 2-cell H, such that (1) J is the sum of an arc apb on that bounding curve of M which contains p, and an arc aqb having only a and b on the boundary of M, and (2) $\overline{H}\cdot s(\overline{H})=0$. Then the transformation s is topological over $\overline{H}$, whence $s(\overline{H})$ is a closed 2-cell having $s(q)$ on its boundary. Let cp be an arc in $H+p$. Then $s(cp)$ is an arc $c'p'$ in $s(H)+p'$ and by Lemma 1, $s(H)$ is in L, hence in K. Suppose that p' is not in K. Then, by Theorem 2, $f(x)\to 0$ as $x\to p'$ on the arc $c'p'$. But $f(x)=f(s(x))$, and as $x\to p'$ on the arc $c'p'$ the point $s(x)\to p$ on the arc cp, and $f(s(x))\to f(p)\neq 0$. This contradiction shows that p' is in K.

LEMMA 7. *If an integral subset of K is the sum of two mutually separated connected sets R_1 and R_2, then either $s(R_1)=R_2$ or else $s(R_1)=R_1$.*

This lemma follows immediately from the facts that the continuous image of a connected set is connected, and that s is of period 2; i.e., that $s(s(A))=A$.

THEOREM 6. *The set K has at most two components. If it has two components R_1 and R_2, then $s(R_1)=R_2$, and R_1 and R_2 have the same boundary.*

Proof. Let R_1 be a component of K. Then $s(R_1)$ is also a component of K. If q is on the boundary of R_1, then there is an arc pq in R_1+q. On this arc $f(x)\to 0$ as $x\to q$ (Theorem 2). Hence $s(pq-q)+q$ is an arc $p'q$ in $s(R_1)+q$, and therefore q is on the boundary of $s(R_1)$. Similarly, every boundary point of $s(R_1)$ is on the boundary of R_1. The proof will be completed by showing that $\overline{R}_1+s(R_1)=M$.

Let $N=\overline{R}_1+s(R_1)$, and suppose that there exists a point t in $M-N$. There is an arc tq having only q in the closed point set N. Clearly q is on the boundary of R_1. Furthermore, q is a limit point of $M-K$, for if it were an isolated point in this set then some point of R_1 could be joined to t by an arc not hitting the boundary of R_1. Then we can apply Lemmas 4 and 5 and get a closed point set J which separates t from q, and such that J is a subset of $R_1+s(R_1)$, except possibly for one or two points on the boundary of R_1. Then $J\cdot tq=0$, since q is the only point of N on tq. This contradicts the fact that J separates t and q.

LEMMA 8. *If p is any point of $M-K$, then for every $\epsilon>0$ there is an open set $W\supset p$ such that*

(1) *$\overline{W}$ is a closed 2-cell of diameter less than ϵ;*

(2) *J, the boundary of W with respect to M, is a simple closed curve or an arc with both end-points on a single bounding curve of M; and*

(3) *$J\cdot(M-K)$ consists of 0, 1, or 2 points.*

If p is not a limit point of $M-K$, then the result is obvious. If p is a limit point of $M-K$, then the proof results from an application of Lemmas 4 and 5. The sets W and J given by Lemma 5 have the desired properties.

THEOREM 7 *Let N be a component of $M-K$. Then N is a point, an arc, or a simple closed curve, and no point of N is a limit point of $M-K-N$.*

Proof. It follows from Lemma 8 that every point of the closed and compact set $M-K$ is of Menger order 0, 1, or 2 with respect to this set. Hence each component of $M-K$ is a point, an arc, or a simple closed curve. Let q be a point of a component N of $M-K$, and suppose that q is a limit point of $M-K-N$. We can apply Lemma 5 and get an open set W, with boundary J, having properties (1), (2), and (3) of Lemma 5, and the property that no arc cc' exists in $K\cdot\overline{W}$, for any point c. We suppose that the set $\overline{W}\cdot(M-K)=T_1+T_2$, mutually separated sets. Then there exists a point u such that u and u' are in K and in $\overline{W}^0$ (the open 2-cell whose closure is $\overline{W}$). Then there exist two arcs aub and $au'b$ such that (1) a and b are in T_1 and T_2, respectively, (2) $s(aub-a-b)=au'b-a-b$, and (3) $(aub+au'b)-a-b$ is in $K\cdot\overline{W}^0$. Then $aub+au'b$ is a simple closed curve J_1 bounding an open 2-cell W_1. Neither of the mutually exclusive closed sets $T_1\cdot\overline{W}_1$ and $T_2\cdot\overline{W}_1$ separates u from u' in $\overline{W}_1$, but their sum does. But this is impossible. Hence $(M-K)\cdot\overline{W}$ is connected. Since it contains q it is in N. Then q is not a limit point of $M-K-N$.

THEOREM 8. *If a component N of $M-K$ is an arc, then each of its end-points, but no other point, is on a boundary curve of M; if N is a point or a simple closed curve, then no point of N is on any boundary curve of M.*

Proof. If p is an end-point of an arc N which is a component of $M-K$, then by Lemma 4 $M-K$ locally separates M at p, and hence by Theorem 7 N locally separates M at p. But this is impossible if p has an open 2-cell neighborhood in M. Hence p is on a bounding curve of M.

Let p be a point of order 2 on some component N of $M-K$, and let apb be an arc which is a subset of N. There is an arc aqb having only a and b in $M-K$ and such that the simple closed curve $aqb+apb$ bounds an open 2-cell U which is in K. Then it follows that $U+s(U)+apb$ contains an open 2-cell containing p.

In a similar way it can be shown that if the point p is a component of $M-K$, then it lies in an open 2-cell.

PART II

LEMMA 1. *If* $\limsup A_n=A$, *then* $\limsup s(A_n)\subset A+s(A)$.

Proof. Let p be a point of $\limsup s(A_n)$. Then there exists a sequence of points $\{p_n\}$, $p_n\in s(A_n)$ $(n=1,2,\cdots)$ such that p is a sequential limit point of $\{p_n\}$. Since G is upper semi-continuous, the sequence $\{s(p_n)\}$, or some sub-

quence, has as limit point either p or $s(p)$. But clearly $s(p_n) \varepsilon A_n$, since $p_n \varepsilon s(A_n)$. Hence if p is a limit point of $\{s(p_n)\}$, then $p \varepsilon \lim\sup A_n$. If $s(p)$ is a limit point of $\{s(p_n)\}$, $s(p) \varepsilon \lim\sup A_n = A$ and $p \varepsilon s(A)$. In either case the lemma is proved.

LEMMA 2. *Given $\epsilon>0$, there exists $\delta>0$ such that if A is a point set of diameter less than δ, then $s(A)$ is the sum of two sets, each of diameter less than ϵ.*

Proof. Suppose the lemma is false. Then there exists an $\epsilon>0$ and a sequence of sets, $\{A_n\}$, such that $d(A_n)<1/n$, but for no n can $s(A_n)$ be expressed as the sum of two sets, each of diameter less than ϵ. Let $\{A_{n_i}\}$ be a subsequence of $\{A_n\}$ such that $\lim\sup A_{n_i} = \lim\inf A_{n_i} = p$, a point. Then by Lemma 1, $\lim\sup s(A_{n_i}) \subset p+s(p)$. But then clearly, for n_i large enough, $s(A_{n_i}) \subset s(p, \epsilon/2)+s(s(p), \epsilon/2)$, and thus $s(A_{n_i})$ is the sum of two sets, each of diameter less than ϵ. This contradiction proves the lemma.

We now have as a corollary of Lemma 2 the following:

LEMMA 3. *Given $\epsilon>0$, there exists $\delta>0$ such that if $d(A)<\delta$ and $s(A)$ is connected, then $d(s(A))<\epsilon$.*

LEMMA 4. *If p is a point of $M-K$ and if q_1p, q_2p, q_3p and q_4p are arcs in $K+p$ with only the point p in common and such that $s(q_1p-p)=q_2p-p$ and $s(q_3p-p)=q_4p-p$, then these arcs have the cyclic order q_1p, q_3p, q_2p, q_4p about p.*

Proof. Suppose the lemma is false. Choose $\epsilon>0$ so that the sphere $S(p, 4\epsilon)$ is an open 2-cell which contains no q_i $(i=1, 2, 3, 4)$ and intersects no component of $M-K$ other than the one to which p belongs. Then for every n we can find a point b_n of q_1p and an arc b_nb_n' in K such that $\rho(b_n, p)<\min(1/n, \epsilon)$ and $d(b_nb_n')<1/n$ and $b_nb_n' \cdot \sum_{i=1}^4 q_ip = b_n+b_n'$.

Now $b_nb_n'+s(b_nb_n')$ is a connected integral set, and hence by Lemma 3, for $n>n_0$, $d(b_nb_n'+s(b_nb_n'))<\epsilon$. Now $b_nb_n'+s(b_nb_n')$ contains a simple closed curve J_n, which is an integral subset of K(9). But since, for $n>n_0$, $J_n \subset S(p, 4\epsilon)$ and $J_n \cdot \sum_{i=1}^4 q_ip \subset b_n+b_n'$, it follows that J_n contains neither p nor any other point of $M-K$ in its interior. Now let R be the component of K which contains J_n, and let I_n be the interior of J_n. Since $s(R)\cdot R\neq 0$, it follows from Theorem 6 of Part I that $s(R)=R$. For n sufficiently large, it is easy to see that $R-(J_n+I_n)$ is connected. Now we cannot have $s(J_n+I_n)=J_n+I_n$, for that would contradict the principal theorem of *Transformations*. Hence by Lemma 7 of Part I, $s(I_n)=R-(J_n+I_n)$. But there is a positive number ϵ_1 such that, for $n>n_1$, $d(R-(J_n+I_n))>\epsilon_1$, and hence, by Lemma 3, a corresponding positive number δ exists such that $d(I_n)>\delta$, for every $n>n_1$. But $\lim d(J_n)=0$, and since M is compact it follows that $\lim d(I_n)=0$. This contradiction proves the lemma.

We are now in a position to prove

(9) See the argument early in §8 of *Transformations* proving the existence of a simple closed curve.

THEOREM 1. *If T is a 2-to-1 transformation defined over M, then M can be so triangulated that the image under s of every n-cell α of the triangulation is an n-cell of the triangulation different from α ($n=0, 1, 2$). Hence $\chi(M)$ is even, and $\chi(T(M))=\chi(M)/2$.*

Proof. It follows from Theorem 5 of Part I that $M-K$ is an integral set. Also, $M-K$ is a compact metric space, and consists of a finite number of simple closed curves $J_1, J_2, \cdots, J_{n_1}$, of isolated points $p_1, p_2, \cdots, p_{n_2}$, and of arcs $v_1, v_2, \cdots, v_{n_3}$ whose end-points lie on the bounding curves of M; and each bounding curve of M contains either two or no points of $M-K$. (This follows easily from Theorems 7 and 8 of Part I.) Moreover, from Theorem 5 of Part I, if p is an isolated point of $M-K$, then $s(p)$ is also an isolated point of $M-K$. Hence T is a 2-to-1 transformation defined over $M-K$, and the point set K_1, over which s is continuous relative to $M-K$, is open and dense in $M-K$. In view of this it can be shown(10) that $K_1 \supset \sum_{i=1}^{n_2} p_i + \sum_{i=1}^{n_3} v_i$, and that J_i contains either no point or exactly two points, u_{1i} and u_{2i}, of $M-K-K_1$.

Now let m be a positive number such that the distance between any two components of $M-K$ is greater than m, and the distance between any two points of $M-K-K_1$ is greater than m. We now choose four positive numbers $\epsilon_1 > \epsilon_2 > \epsilon_3 > \epsilon_4$ with the following properties:

(1) $4\epsilon_1 < m$;

(2) if B and $s(B)$ are connected sets and if $d(B) < \epsilon_2$, then $d(s(B)) < \epsilon_1$ (see Lemma 3);

(3) any simple closed curve of M of diameter less than ϵ_3 bounds a 2-cell of M of diameter less than ϵ_2(11); and

(4) if A and $s(A)$ are connected sets, and if $d(A) < \epsilon_4$, then $d(s(A)) < \epsilon_3$ (see Lemma 3).

Now consider an isolated point p_j of $M-K$. Let q_1 be a point of K such that $\rho(q_1, p_j) < \epsilon_4$ and let γ_1 be an arc of diameter less than ϵ_4 from q_1 to p_j. Let γ_2 be the arc $p_j + s(\gamma_1 - p_j)$. Let q_3 be a point of K such that $\rho(p_j, q_3) < \epsilon_4$, and $q_3 \not\in \gamma_1 + \gamma_2$. Let γ_3 be an arc of diameter less than ϵ_4 from q_3 to p_j such that $\gamma_3 \cdot (\gamma_1 + \gamma_2) = p_j$. Let γ_4 be the arc $p_j + s(\gamma_3 - p_j)$. Now by Lemma 4 these arcs have the cyclic order, $\gamma_1, \gamma_3, \gamma_2, \gamma_4$, about p. We now find two points, $r_1 \in \gamma_1$ and $r_3 \in \gamma_3$, and two arcs, β_1 from r_1 to r_3, and β_3 from r_3 to r_1' with the following properties:

(1) $(\sum_{i=1}^{4} \gamma_i)(\beta_1 + \beta_3 + \beta_1' + \beta_3') = r_1 + r_3 + r_1' + r_3'$; and

(2) if $r_1' = r_2$ and $r_3' = r_4$, and if δ_i denotes the subarc of γ_i from p_j to r_i ($i=1, 2, 3, 4$), then $d(\beta_1 + \delta_1 + \delta_3) < \epsilon_4$ and $d(\beta_3 + \delta_3 + \delta_2) < \epsilon_4$. It follows from the definition of ϵ_3 and the fact that $\epsilon_4 < \epsilon_3$, that $\beta_1 + \delta_1 + \delta_3$ and $\beta_3 + \delta_3 + \delta_2$ are simple closed curves which are the boundaries of closed 2-cells, λ_1 and λ_3,

(10) K_1 will be used to denote this set throughout the rest of the paper. The proofs of the statements of this sentence, while not trivial, are sufficiently straightforward to be omitted.

(11) It is easy to see that, since M is compact, ϵ_3 can be chosen to satisfy this property.

respectively. It is easy to see that $s(\beta_1)$ is an arc β_2 from r_2 to r_4, and $s(\beta_3)$ is an arc β_4 from r_4 to r_1; and further, that $s(\lambda_1^0)$ is an open 2-cell λ_2^0 bounded by $\beta_2+\delta_2+\delta_4$, and $s(\lambda_3^0)$ is an open 2-cell λ_4^0 bounded by $\beta_4+\delta_4+\delta_1$. Thus the neighborhood of p_j has been triangulated in accordance with the theorem.

We suppose that this has been done for every p_j $(j=1, 2, \cdots, n_1)$. In a similar manner we triangulate the neighborhood of each of the simple closed curves $J_1, J_2, \cdots, J_{n_1}$ and each of the arcs $v_1, v_2, \cdots, v_{n_2}$ in such a way that these simple closed curves and arcs appear in the triangulation as sums of 1-cells and vertices of the triangulation which map under s into 1-cells and vertices of the triangulation. We take care, also, to make every point of $M-K-K_1$ a vertex of the triangulation.

We now let A_1 denote the complex which is composed of all these neighborhoods of the components of $M-K$ so triangulated, and let $H=M-A_1^*$. Then $\overline{H}$ is a closed and compact point set over which f is continuous and positive. Hence there is a positive number ψ such that $f(x)>2\psi$ if $x \in \overline{H}$. Now we choose three positive numbers ψ_1, ψ_2, and ψ_3 as follows: (1) if γ is a simple closed curve of diameter less than $4\psi_1$, then γ bounds a 2-cell of diameter less than ψ (this implies $4\psi_1<\psi$); and (2) if B and $s(B)$ are connected sets, then if $d(B)<\psi_2$, it follows that $d(s(B))<\psi_1$; and if $d(B)<\psi_3$, it follows that $d(s(B))<\psi_2$. (See Lemma 3. These conditions imply $\psi_1>\psi_2>\psi_3$.) If A_1 contains a 1-cell α such that $d(\alpha)\geqq\psi_3$, let A_2 be a subdivision of A_1 containing no such 1-cell but still having the property that the image under s of an n-cell of A_2 is an n-cell of A_2 $(n=0, 1, 2)$. If A_1 contains no such 1-cell, $A_2=A_1$.

Now it may be possible to find either one, two, or three arcs whose end-points are vertices of A_2 but which otherwise lie in $M-A_2^*$ and which, together with two, one, or no arcs, respectively, of A_2^*, bound a 2-cell ϕ of diameter less than ψ whose interior lies in $M-A_2^*$, but not in $S(A_2^*, \psi_3)$. If such a possibility exists, we add to A_2 this 2-cell ϕ and also $s(\phi)$. Since $d(\phi)<\psi$, $\phi\cdot s(\phi)=0$. After extending A_2 in this manner as many times as possible, successively, we call the extended complex A_3.

If $M-A_3^*$ contains an open 2-cell α^0 such that the boundary of α^0 consists of three vertices and three 1-cells of A_3 and $s(\alpha^0)\cdot\alpha^0=0$, then we add α^0 and $s(\alpha^0)$ to A_3. After adding all such 2-cells to A_3, we call the new complex A_4. We then obtain A_5 from A_4 by subdivision, in the same way that we obtained A_2 from A_1.

Now let $r_1, r_2, \cdots, r_k$ denote the vertices of A_5 which are on the boundary of $M-A_5^*$. Let r_{k+1} be a point of $M-A_5^*$ for which $\psi_3<\rho(r_{k+1}, A_5^*)<\psi_2$. Join r_{k+1} to two points, r_{i_2} and r_{i_3}, which are end-points of the same 1-cell α_1 of A_5 by arcs α_2 and α_3 in such a way that (1) $\alpha_2\cdot\alpha_3=r_{k+1}$, (2) $\alpha_2\cdot A_5^*=r_{i_2}$ and $\alpha_3\cdot A_5^*=r_{i_3}$, and (3) $d(\alpha_i)<2\psi_1$ $(i=1, 2)$. To see that this is possible, we draw an arc β of diameter less than ψ_2(12), from r_{k+1} to some point p in the interior

(12) This is possible since the metric ρ which we are using has the property that if H is closed and $\rho(x, H)<\epsilon$, then there exists an arc from x to a point of H of diameter less than ϵ.

of a 1-cell α_1 on the boundary of A_5^*. Since no 1-cell of A_5 has diameter as great as ψ_3, we can draw an arc in $M-A_5^*$ from r_{k+1}, running along very close to β and then running along close to α_1 until we get near an end-point r_{i_2} of α_1. Then we run our arc into r_{i_2} and call the arc α_2. Similarly, on the other side of β, we draw α_3 from r_{k+1} to r_{i_3}. If in drawing these arcs we stay close enough to β and then to α_1, neither α_2 nor α_3 can have diameter greater than $\psi_2+\psi_3$, which is less than $2\psi_1$. Moreover, $\alpha_1+\alpha_2+\alpha_3$ is a simple closed curve of diameter less than $4\psi_1$, and hence by definition of ψ_1 bounds a 2-cell η of diameter less than ψ. It follows that $s(\eta)$ is a 2-cell such that $\eta\cdot s(\eta)=0$ and $s(\eta)\cdot A_5=s(\alpha_1)$. We add η and $s(\eta)$ to A_5 and call the resulting complex A_6.

We now begin over again, obtaining A_7 from A_6 in the same way that we obtained A_2 from A_1, etc. It is clear that the method of extension from A_5 to A_6 can be carried out only a finite number of times. For otherwise M would contain an infinite point set with no limit point, since we require that the new vertex have a distance greater than ψ_3 from the point set covered by the complex. But when this method of extension cannot be repeated that means that the complex we have, A_n, has the property that there is no point of $M-A_n^*$ which has a distance greater than ψ_3 from A_n^*. When this is the case, it is easy to see that a finite number of applications of the methods of extension from A_2 to A_3 and from A_3 to A_4 will give a complex which covers M. Hence the theorem is proved.

Lemma 5. *Let p be an interior point of M. If p is neither an isolated point of $M-K$, nor a point of $M-K-K_1$, then $T(p)$ has a 2-cell neighborhood in $T(M)$. Otherwise $T(p)$ has a neighborhood in $T(M)$ which is homeomorphic to a neighborhood of the vertex of the double cone $x^2=y^2+z^2$.*

Hereafter when we speak of a manifold with identifications we shall always mean by "identifications" a finite number of points with neighborhoods homeomorphic to a neighborhood of the vertex of the double cone.

Proof. If p belongs to K, the result follows immediately; for s is locally a homeomorphism at every point of K, and if U is an open subset of K such that $U\cdot s(U)=0$, then $s(U)$ and $T(U)$ are homeomorphic. But this means that $T(U)$ contains an open set which is an open 2-cell containing $T(p)$.

If p belongs to $(M-K)\cdot K_1$, but is not an isolated point of $M-K$, then p belongs to an arc or a simple closed curve of $M-K$. Let c denote the arc or the simple closed curve. By triangulating the neighborhood of p and of $s(p)$ in the same way in which they were triangulated in the proof of Theorem 1, and taking care to have neither p nor $s(p)$ be a vertex in this triangulation, we obtain mutually exclusive open 2-cells V and W containing p and $s(p)$, respectively, and having the following properties:

(1) $V-c=V_1+V_2$, mutually separated open 2-cells such that $s(V_1)=V_2$;
(2) $V\cdot c$ is an open 1-cell lying in K_1;
(3) $V\cdot(M-K)=V\cdot c$;

(4) $W\cdot s(V\cdot c)=W\cdot(M-K)$;

(5) $W-s(V\cdot c)=W_1+W_2$, mutually separated open 2-cells such that $s(W_1)=W_2$.

It follows that $s(V\cdot c)$ is an open 1-cell, and that $T(V\cdot c)=T(s(V\cdot c))$ is an open 1-cell. Likewise, each of the sets $T(V)$ and $T(W)$ is homeomorphic to the intersection of the interior of the unit sphere in the Euclidean plane with the half-plane $x\geqq 0$. Moreover, $T(V)\cdot T(W)=T(V\cdot c)$, an open 1-cell. Hence $T(V+W)$ is an open 2-cell containing $T(p)$. Furthermore, it is clear that $T(V+W)$ is an open set in $T(M)$, since $V+W$ is an integral open set in M.

If p is an isolated point of $M-K$, then, as we remarked in the proof of Theorem 1, $s(p)$ is also. We triangulate the neighborhood of p and of $s(p)$ as we did in the proof of Theorem 1. From the description of that triangulation, it is easy to verify that if $\pi(p)$ denotes the open 2-cell containing p whose closure is $\lambda_1+\lambda_2+\lambda_3+\lambda_4$ (see the proof of Theorem 1 for the meaning of λ_i), then $T(\pi(p))$ is an open 2-cell. Similarly for $\pi(p')$. But $T(\pi(p))\cdot T(\pi(p'))=T(p)$. Hence $T(\pi(p)+\pi(p'))$ is homeomorphic to a neighborhood of the vertex of the double cone $x^2=y^2+z^2$, and since $\pi(p)+\pi(p')$ is an integral open set in M, $T(\pi(p)+\pi(p'))$ is open in $T(M)$, and the lemma follows for this case.

The case in which p belongs to $M-K-K_1$ is handled in a somewhat similar manner.

LEMMA 6. *Let p be on one of the boundary curves of M. Then $T(p)$ has a neighborhood in $T(M)$ which is homeomorphic to a neighborhood of p in M.*

The proof follows in much the same way that the proof of the preceding lemma followed.

THEOREM 2. *If M is a compact manifold, then $T(M)$ is a compact manifold or can be obtained from a compact manifold by the identification by pairs of a finite number of points. If M is a compact manifold with boundary, then $T(M)$ is a compact manifold with boundary or can be obtained from a compact manifold with boundary by the identification by pairs of a finite number of interior points.*

Proof. The first statement follows as a corollary of Lemma 5. To prove the second statement, we first show that if c is a boundary curve of M, then $T(c)$ is a simple closed curve. For suppose first that $K\not\supset c$. Then by Theorem 8 of Part I, $c\cdot K\neq 0$. Let p be a point of $c\cdot K$, and let pq be a subarc of c which lies in $K+q$ but not in K. Then $p'q=s(pq-p)+q$ is a subarc of c, by virtue of Theorem 2 of Part I and the fact that the image under s of a boundary point of M is a boundary point of M. Let pr be a subarc of c such that $pr\cdot pq=p$ and $K+r\supset pr$ but $K\not\supset pr$. (It is easy to see that there must be a point $r\neq q$ such that $r\in c\cdot(M-K)$. For otherwise, as a variable point x

moved continuously along $c-pq-p'q+p$ from p, $s(x)$ would move continuously along $c-pq-p'q+p'$ and a two-to-one transformation would be defined on the arc $c-pq-p'q+p+p'$, in contradiction to the result of O. G. Harrold([1]).) Then as above, $p'r=s(pr-r)+r$ is a subarc of c. Then clearly $c=pq+p'q+pr+p'r$, and a direct argument shows that $s(q)=r$. Hence in this case $T(c)$ is a simple closed curve.

Now suppose that $K\supset c$. Then $s(c)$ is a simple closed curve which is a boundary curve of M, and hence either $s(c)=c$ or else $s(c)\cdot c=0$. In either case $T(c)$ is a simple closed curve.

Now by combining the fact that $T(c)$ is a simple closed curve with Lemma 5 and Lemma 6, we see that the second statement of our theorem is proved.

Moreover, it is easy to see that we have also proved

LEMMA 7. *If M has n boundary curves, then $T(M)$ has k boundary curves, where $n/2\leqq k\leqq n$.*

We now prove

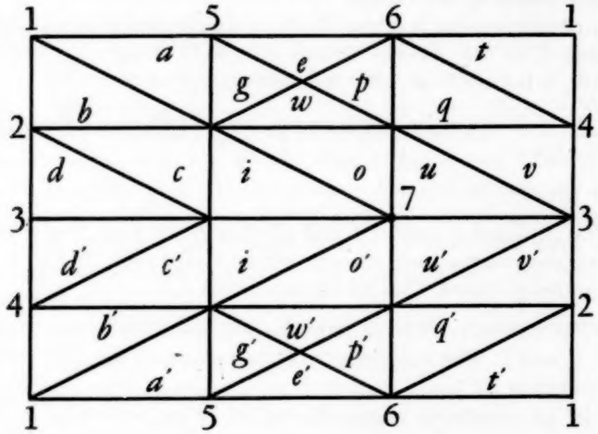
THEOREM 3. *Given a space M and a space B_k, a necessary and sufficient condition that there exist a 2-to-1 transformation T such that $T(M)=B_k$ is that M and B_k be properly related.* (*See the introduction.*)

Proof. The necessity follows immediately from Theorems 1 and 2 and Lemmas 5, 6, and 7. The sufficiency will be proved by actually constructing the transformation T.

Let M be an orientable manifold, and let M be embedded in Euclidean 3-space in such a way that M is symmetric with respect to the xy-plane, and the common part of M and the xy-plane consists of h simple closed curves $c_1, c_2, \cdots, c_h$ if $\chi(M)=2(h-2)$. If p is a point of M not in the xy-plane, we define $s(p)$ as the reflection of p in the xy-plane. If B_k is an orientable manifold, then h is even and we let s map c_1 into c_2, c_3 into $c_4, \cdots, c_{h-1}$ into c_h topologically. When s is defined, T is determined, and the theorem is proved for this case. If B_k is a non-orientable manifold, we define s exactly as before except that s maps c_h into itself by identifying diametrically opposite points, in case h is odd, and it maps both c_h and c_{h-1} into themselves in this manner if h is even. If B_k is an orientable manifold with b identifications, then $h-b$ is even and non-negative. For if a manifold N has $\chi(N)=v$, and if N_1 is obtained from N by identifying b pairs of points, then $\chi(N_1)=v+b$. Hence we define $s(p)$ as before for a point p not in the xy-plane, and we let x map c_i $(i=1, 2, \cdots, b)$ into itself by a reflection in a diameter and the identification of the two points of c_i which lie on the diameter. And we let s map c_{b+1} into c_{b+2}, c_{b+3} into $c_{b+4}, \cdots, c_{h-1}$ into c_h, topologically. If B_k is a non-orientable manifold with b identifications, s is defined as in the preceding case except that if $h-b$ is odd, then s maps c_h into itself by identifying diametrically

opposite points, and if $h-b$ is even then s maps both c_{h-1} and c_h into themselves in that manner. This completes the cases in which M is an orientable manifold.

If M is a non-orientable manifold, and $\chi(M)$ is even, then either M is a Klein's bottle, in which case $\chi(M)=0$, or else M can be obtained from a Klein's bottle by inserting h handles, in which case $\chi(M)=2h$. If M is a Klein's bottle, let M be triangulated as in the figure. The numbers in this figure refer to the vertices by which they are placed and the letters denote the



2-cells in which they are placed. The top edge of the figure is identificed with the bottom edge, and the points of the left edge reading up are identified with those of the right edge reading down, as the numbers indicate. Let s map the 2-cells above the horizontal bisector 373 into those below 373 by reflection in 373, as the primes indicate. This defines s for every point except those points of the horizontal bisector 373 and of the edge 1561; we call these two simple closed curves c_1 and c_2. If B_k is a torus, we let s map c_1 into c_2 topologically; if B_k is a projective plane with one identification we let s map c_1 into itself by identifying diametrically opposite points and c_2 into itself by identifying points by reflection in a diameter and identifying the points of c_2 on the diameter; if B_k is a Klein's bottle, we let s map c_i into itself by identification of diametrically opposite points $(i=1, 2)$; if B_k is a sphere with two identifications, we let s map c_i into itself by reflection in a diameter and identification of the points of c_i on the diameter $(i=1, 2)$.

If M is non-orientable and $\chi(M)=2h>0$, then M can be constructed as follows: Let N_1 be the manifold with boundary obtained from the figure by deleting the open 2-cells w and w', and let s map $N_1-(c_1+c_2)$ into itself in the manner described above. Let N_2 be an orientable manifold for which

$\chi(N_2)=2h-4$. Let N_2 be embedded in Euclidean 3-space so that it is symmetric with respect to the xy-plane and the common part of N_2 and the xy-plane consists of h simple closed curves. Let s map the points of N_2 above the xy-plane into those below the xy-plane by reflection in that plane. Let α be an open 2-cell in N_2 lying above the xy-plane and having a simple closed curve as boundary, and let $\alpha'=s(\alpha)$. Delete α and α', and identify the boundaries of α and α' with the boundaries of w and w', respectively. Then we have defined s over M except for $h+2$ simple closed curves. We define s over these simple closed curves, using one of, or a combination of, the three methods already described for defining s over simple closed curves, depending on the character of the image space B_k.

The definition of T in the cases in which M is a manifold with bounding curves is analogous to its definition in the cases already treated.

Conclusion. It is known that in some cases a space M may be mapped into a space B in a continuous, exactly 2-to-1 fashion, in at least two essentially different ways. For example, a sphere may be mapped into a projective plane (1) by identifying diametrically opposite points, or (2) by identifying point pairs which are symmetrical with respect to the equatorial plane, and then identifying diametrically opposite points on the equator. It is easy to show that the two mappings so defined are not topologically equivalent([13]). The following problem naturally arises: *For a given M and B how many topologically different 2-to-1 continuous mappings of M into B are there?* It seems very likely that this number is finite.

([13]) For a definition of this term see G. T. Whyburn, *Interior transformations on compact sets*, Duke Mathematical Journal, vol. 3 (1937), p. 373, footnote 8.

Duke University,
Durham, N. C.

THE INTEGRAL REPRESENTATION OF WEAKLY ALMOST-PERIODIC TRANSFORMATIONS IN REFLEXIVE VECTOR SPACES

BY

EDGAR R. LORCH

I. Introduction

The first analysis of the structure of transformations of a type somewhat similar to that treated below was given by D. Hilbert in his memoir on bounded quadratic forms in infinitely many variables (1906). In the language freely used at present, Hilbert established that if H is a bounded symmetric transformation in the (Hilbert) space of vectors $x=(x_1, x_2, \cdots)$ with $\sum |x_\alpha|^2 < \infty$, then H may be represented in the form

$$H = \int \lambda dE(\lambda). \tag{1}$$

The $E(\lambda)$ are orthogonal projections, that is to say, transformations which are the identity transformation on a closed linear manifold in the space and which are the zero transformation on the orthogonally complementary manifold. As λ varies increasingly from some bound $-M$ to M, $E(\lambda)$ sweeps monotonously from the zero to the identity transformation. The values λ, $-M \leq \lambda \leq M$, are classed as being regular or belonging to the "spectrum" of H according to the behaviour of $E(\lambda)$. The complete analysis of the spectrum of H leading to the integral representation (1) is sometimes described as establishing the spectral resolution of H. The formula (1) is the Hilbert space correspondent of the fundamental theorem of matrix theory to the effect that every symmetric $n \times n$ matrix may be reduced to diagonal form by a proper choice of axes.

Since the publication of Hilbert's result, the theory initiated by him has been elaborated in many directions. But the class of transformations in infinite dimensional spaces whose structure is thoroughly known has not grown. There is at least one important exception: the restriction of boundedness upon H has been deleted. It is known as a result of the work of Carleman, von Neumann, F. Riesz, Stone, and Wintner(1), that the equation (1) sub-

Presented to the Society in two parts, the first under the title *Spectral analysis of weakly almost periodic transformations in reflexive vector spaces* on December 29, 1939; the second, under the present title on October 26, 1940; received by the editors May 8, 1940. Abstracted in the Bulletin of the Society, 45-11-401 and 46-7-361.

(1) Carleman, *Sur les Équations Intégrales Singulières à Noyau Réel et Symétrique*, Uppsala, 1923; J. von Neumann, Mathematische Annalen, vol. 102 (1929), pp. 49–131; F. Riesz, Acta Litterarum ac Scientiarum (Szeged), vol. 5 (1930), pp. 19–54; Stone, *Linear Transforma-*

sists in the non-bounded case. Here the range of integration necessarily is not bounded on the λ axis. Suggestive of a new point of departure is a result established by von Neumann[2] that a rotation or unitary transformation V in Hilbert space has a "diagonal" representation of the form

$$V = \int e^{i\lambda} dE(\lambda). \tag{2}$$

Since simple correspondences may be established between unitary and symmetric (or better, self-adjoint) transformations, it is correct to state that the only transformations analyzed to the present (leaving aside certain considerably simpler types such as the completely continuous transformations) are the rotations in Hilbert space and the functions which they beget, e.g., $H = -i(V-I)(V+I)^{-1}$, e^H, etc.

A study of this theory reveals that its development is wedded to and completely dominated by the concept of orthogonality. It is the nature of this concept which renders impossible the mere extension of the notion of symmetric transformation to non-Hilbert spaces. Properly speaking, orthogonality within a space $\mathfrak{B}$ is meaningless; it is biorthogonality between $\mathfrak{B}$ and its adjoint space $(\mathfrak{B})$ which is significant. As Hilbert space is distinguished by the equation $\mathfrak{B} = (\mathfrak{B})$, biorthogonality is operative within the space itself and is then called orthogonality. Lacking a theory of orthogonality, entirely new methods have to be developed if one is to progress in the study, for example, of the structure of rotations in various non-Hilbert spaces. In the subsequent pages, the spectral resolution of a type of transformation in certain rather general spaces is obtained. All rotations fall under this type. But even in Hilbert space, new classes are analyzed for the first time.

The nature of the space for which the results are valid will first be described. For an arbitrary normed linear vector space $\mathfrak{B}$, the inclusion $((\mathfrak{B})) \supset \mathfrak{B}$ is easy to establish. A space distinguished by the relation

$$\mathfrak{B} = ((\mathfrak{B})) \tag{3}$$

is called *reflexive* by the author. It is for reflexive spaces that a spectral theory is developed.

The precise nature of the transformation V for which formula (2) may be established will be described briefly. Starting with a V such that V^{-1} exists (and is bounded), the set $\{V^n f\}$, $n = 0, \pm 1, \pm 2, \cdots$, f fixed in $\mathfrak{B}$, is examined. Borrowing ideas from the theory of almost-periodic functions, one defines V to be weakly almost-periodic if the set $\{V^n f\}$ is weakly conditionally

tions in Hilbert Space and Their Applications to Analysis, American Mathematical Society Colloquium Publications, vol. 15, New York, 1932, chap. 5; Wintner, *Spektraltheorie der unendlichen Matrizen*, Leipzig, 1929, chap. 6.

(2) J. von Neumann, op. cit., p. 119.

compact for all $f \varepsilon \mathfrak{B}$. That is, if relative to the group V^n, all $f \varepsilon \mathfrak{B}$ are almost-periodic, then V is called almost-periodic—all this in the weak sense. The almost-periodicity of V is characterized by the theorem: *V is weakly almost-periodic if and only if it is uniformly bounded,* $\|V^n f\| \leq K\|f\|$, $n=0, \pm 1, \pm 2, \cdots$, *$f$ arbitrary.* It is these transformations which are analyzed.

The results of the analysis yield a structure $\{\mathfrak{E}_\lambda, \mathfrak{F}_\lambda\}$, $-\infty < \lambda < \infty$, strongly resembling a resolution of the identity and replacing it in the present instance. The $\mathfrak{E}_\lambda$ and $\mathfrak{F}_\lambda$ are closed linear manifolds having only the origin in common and together spanning $\mathfrak{B}$. Further, $\mathfrak{E}_\lambda$ ($\mathfrak{F}_\lambda$) is a monotone function of λ increasing (decreasing) from 0 ($\mathfrak{B}$) to $\mathfrak{B}$ (0). The manifold pairs reduce V. Whether $\mathfrak{E}_\lambda$ and $\mathfrak{F}_\lambda$ are disjoint (see below) is yet to be determined. The reconstruction of V, in a sense justifying the use of (2), from the manifolds which it determines is carried out in Theorem 8. The spectral character of V is completely determined by the structure of $\{\mathfrak{E}_\lambda, \mathfrak{F}_\lambda\}$—Theorem 9. As an instance, exp $(i\lambda)$, $-\pi < \lambda < \pi$, is in the resolvent set of V if and only if λ is a point of constancy of $\{\mathfrak{E}_\lambda, \mathfrak{F}_\lambda\}$.

In the final chapter, the transformation $H = -i(V-I)(V+I)^{-1}$ is subjected to a thorough analysis. Formula (1) is established for it (Theorem 10), and it is shown that its spectrum may be read accurately in $\{\mathfrak{E}_\lambda, \mathfrak{F}_\lambda\}$—Theorem 11.

The present paper is based at least in part upon two previous papers of the author. The reader is referred to these at various points in the development. One of these papers, on the so-called mean-ergodic theorem in reflexive vector spaces, embodies the earliest results obtained in the present investigation, and quite naturally was to have been incorporated in the present paper (Theorems 3 and 4). However, the interest which similar theorems had aroused more recently made it desirable to publish the result separately (see footnote 7). In addition to these sources, no acquaintance with linear transformation theory is necessary to an understanding of the contents, other than may be found in the elementary chapters of Banach's treatise on linear operations.

II. Weakly almost-periodic transformations

The underlying space $\mathfrak{B}$ is assumed to be reflexive throughout (exceptions: Theorems 2 and 4′). That is to say, $\mathfrak{B}$ is a linear normed complex vector space satisfying the identity $\mathfrak{B} = ((\mathfrak{B}))$. Here $(\mathfrak{B})$ denotes the space of all bounded linear functionals defined on $\mathfrak{B}$; $(\mathfrak{B})$ is called the space adjoint to $\mathfrak{B}$. Thus $((\mathfrak{B}))$ is the space adjoint to $(\mathfrak{B})$. Elements in $\mathfrak{B}$ will be denoted by $f, g, h, f_n, \eta, \zeta$, etc.; real numbers by λ, μ, ϵ, etc.; complex numbers by α, β, etc. Transformations will be denoted by T, V, A, B, P, I (the identity), and 0 (the zero), etc. They are linear or distributive, that is, $T(\alpha f + \beta g) = \alpha Tf + \beta Tg$. The bound of a transformation T (if it exists) is denoted by $|T|$, also by K, M, etc. If T is bounded and possesses a bounded inverse T^{-1} ($TT^{-1} = T^{-1}T = I$), T is said to be *bicontinuous*.

If T is bicontinuous, it generates a group of transformations, viz., T^n, $n=0, \pm 1, \pm 2, \cdots (T^0=I)$, with $T^n \cdot T^m = T^{n+m}$. A notion of the theory of almost-periodic functions suggests an analogue for the present study. A function $f(x)$ defined over a group is called almost-periodic if the set of functions $\{f(cx)\}$ with c a parameter ranging over the group is conditionally compact. Consider now the set $\{T^n f\}$, $n=0, \pm 1, \pm 2, \cdots, f \in \mathfrak{B}$ fixed; whether the set is conditionally compact or not, that is, whether a converging subsequence can be drawn from any sequence in $\{T^n f\}$ depends on the type of convergence in question. Of the two types which instantly present themselves, *strong* and *weak* convergence, the latter seems much more fitting here. The significance of the definition which follows is now clear.

DEFINITION. *Let T be a bicontinuous transformation in a reflexive space $\mathfrak{B}$. If the set $\{T^n f\}$, $n=0, \pm 1, \pm 2, \cdots, f$ fixed, is weakly conditionally compact for every $f \in \mathfrak{B}$, T is said to be weakly almost-periodic.*

A space is called weakly compact if every bounded set in the space is weakly conditionally compact. A space is called weakly complete if every weakly converging sequence converges weakly to an element of the space. It is known that reflexive spaces are both weakly compact and weakly complete([3]). Since any weakly converging sequence is necessarily bounded([4]) it follows that in order that T be weakly almost-periodic (for short, w.a.p.), it is necessary and sufficient that for each f the set $\{T^n f\}$ be bounded. Thus for every $f \in \mathfrak{B}$, there must exist a constant K_f such that $\|T^n f\| \leqq K_f \|f\|$, $n=0, \pm 1, \pm 2, \cdots$.

A bicontinuous transformation T is said to be *uniformly bounded* if the bound of T^n satisfies the inequality $|T^n| \leqq K$, $n=0, \pm 1, \pm 2, \cdots$, for some constant K. The first theorem links the concepts of uniform boundedness and weak almost-periodicity.

THEOREM 1. *In a reflexive space, the concepts of uniform boundedness and weak almost-periodicity are equivalent.*

Let T be uniformly bounded, $|T^n| \leqq K$, $n=0, \pm 1, \pm 2, \cdots$. Then for an arbitrary $f \in \mathfrak{B}$, $\|T^n f\| \leqq K\|f\|$. Thus by the argument just given above, T is w.a.p.

Now let T be w.a.p. Then for $f \in \mathfrak{B}$, there exists a constant K_f such that $\|T^n f\| \leqq K_f \|f\|$. The existence of a single constant K valid for all $f \in \mathfrak{B}$ will be established.

Let there be constructed a normed linear complex vector space $\mathfrak{C}$ in the following fashion: The elements of $\mathfrak{C}$ are the objects

$$[f] = \{ \cdots, T^{-2}f, T^{-1}f, f, Tf, T^2 f, \cdots \},$$

([3]) Pettis, *A note on regular spaces*, Bulletin of the American Mathematical Society, vol. 44 (1938), pp. 420–428.

([4]) Banach, *Théorie des Opérations Linéaires*, Warsaw, 1932, p. 133.

$f \varepsilon \mathfrak{B}$. Multiplication by a complex scalar α is defined by $\alpha[f]=[\alpha f]$. Addition of vectors is defined by $[f]+[g]=[f+g]$. The norm of the vector $[f]$ is defined by

$$\| [f] \| = \text{l.u.b.} \|T^n f\|, \qquad n = 0, \pm 1, \pm 2, \cdots .$$

Clearly $\| [f] \| \leqq K_f \|f\|$.

The space $\mathfrak{C}$ is complete. For let $\{ [f]_r \} = \{ [f_r] \}$ be a sequence converging in the above norm,

$$\lim_{r,s\to\infty} \| [f_r] - [f_s] \| = \lim_{r,s\to\infty} \| [f_r - f_s] \| = 0.$$

Thus $\lim_{r,s\to\infty} \|T^n f_r - T^n f_s\| = 0$ uniformly in n. For $n=0$, it is seen that the sequence $\{f_r\}$ converges to some element $f \varepsilon \mathfrak{B}$. Also the convergence of $\{T^n f_r\}$ to $T^n f$ is uniform in n. It is now clear that $\{ [f_r] \}$ converges to $[f]$. For

$$\| [f - f_r] \| = \text{l.u.b.}_n \|T^n(f - f_r)\| \to 0$$

with $r \to \infty$. Thus $\mathfrak{C}$ is complete. Since

$$\text{l.u.b.}_n \|T^n(f+g)\| \leqq \text{l.u.b.}_n \|T^n f\| + \text{l.u.b.}_n \|T^n g\|,$$

$$\| [f + g] \| \leqq \| [f] \| + \| [g] \|.$$

Also $\| [\alpha f] \| = |\alpha| \cdot \| [f] \|$. Hence $\mathfrak{C}$ is a normed linear complex vector space.

Now let U be the mapping of the space $\mathfrak{B}$ upon the space $\mathfrak{C}$ which carries $f \varepsilon \mathfrak{B}$ into $[f] \varepsilon \mathfrak{C}$,

$$Uf = [f].$$

U is distributive. Furthermore, U is closed, that is, if $f_r \to f$ and $[f_r] \to [g]$, then $Uf = [g]$. For since the sequence $\{ [f_r] \}$ converges (which in itself implies that the sequence $\{f_r\}$ converges) and since $f_r \to f$, then $g = f$. This implies the closure of U.

It is known that a closed distributive transformation whose domain is the entire space $\mathfrak{B}$ is bounded(6). Thus $|U| \leqq K$ and

$$\|T^n f\| \leqq \| [f] \| = \|Uf\| \leqq K \|f\|. \tag{4}$$

This proves the theorem.

It is to be noted that all rotations, that is to say, all bicontinuous transformations V which are isometric, $\|Vf\| = \|f\|$, are w.a.p. So are also all periodic transformations. The structure of periodic linear transformations is particularly simple and is given in the next theorem. The results presage those attending the discussion of a general w.a.p. transformation.

(6) Banach, p. 41.

THEOREM 2. *Let T be a bounded linear periodic transformation of period n, $T^n=I$. Let $|T^r|\leqq K$, $r=1, 2, \cdots, n$. Then there exist projections P_r, $r=1, 2, \cdots, n$, with the properties:*

(a) $|P_r|\leqq K$; $P_rP_s=P_sP_r=0$, $r\neq s$.

(b) $P_1+P_2+\cdots+P_n=I$.

(c) $TP_rf=\alpha^rP_rf$ *where* $\alpha=e^{2\pi i/n}$.

Note that $f=f_1+f_2+\cdots+f_n$ provided that

$$f_r = \frac{1}{n}(f+\alpha^{-r}Tf+\alpha^{-2r}T^2f+\cdots+\alpha^{-(n-1)r}T^{n-1}f),$$

$r=1, 2, \cdots, n$. Clearly

$$Tf_r = \alpha^r f_r, \qquad r = 1, 2, \cdots, n.$$

The transformations $f\rightarrow f_r$, $P_rf=f_r$, are distributive.

$$\|f_r\| \leqq \frac{1}{n}(\|f\|+\|\alpha^{-r}Tf\|+\cdots+\|\alpha^{-(n-1)r}T^{n-1}f\|) \leqq K\|f\|,$$

$|P_r|\leqq K$. That $P_r^2f=P_rf_r=f_r=P_rf$ may be ascertained directly from the definition of f_r. Thus the P_r are projections. Similarly it may be seen that $P_rP_s=P_sP_r=0$ if $r\neq s$. Finally, that $P_1+P_2+\cdots+P_n=I$ is the opening statement of this proof.

III. THE REDUCIBILITY OF WEAKLY ALMOST-PERIODIC TRANSFORMATIONS

In the subsequent sections, a w.a.p. transformation will frequently be denoted by V. The spectral analysis of V is based upon the following considerations. Let us assume that a representation (2) is possible for V. Let us grant that the "resolution of the identity" $E(\lambda)$ has a sufficiently smooth structure so that an operational calculus can be based upon it[6]. Then in this calculus, V corresponds to the function $\exp(i\lambda)$. This suggests that the structure of V may be analyzed by means of Fourier series. Of course not any Fourier series is acceptable, but since $|V^n|\leqq K$, any series for which the series of coefficients is absolutely convergent is meaningful and may be useful in the present situation. A series of the type which "splits" V will now be introduced.

Let $f(x)$ be a function of period 2π defined by

$$f(x) = 0, \qquad -\pi \leqq x \leqq 0;$$

$$f(x) = \sin x, \qquad 0 < x < \pi.$$

Let $g(x)$ be defined by $g(x)=\sin x-f(x)$. The Fourier expansion of $f(x)$ is

[6] Lorch, *On a calculus of operators in reflexive vector spaces*, these Transactions, vol. 45 (1939), pp. 217–234. See in particular Definition 4.

$$f(x) = \frac{1}{2}\sin x + \frac{2}{\pi}\left(\frac{1}{2} - \frac{\cos 2x}{1\cdot 3} - \frac{\cos 4x}{3\cdot 5} - \frac{\cos 6x}{5\cdot 7} - \cdots\right)$$

$$(5) \qquad = \frac{1}{4i}(e^{ix} - e^{-ix})$$

$$+ \frac{1}{\pi}\left(1 - \frac{1}{1\cdot 3}(e^{2ix} + e^{-2ix}) - \frac{1}{3\cdot 5}(e^{4ix} + e^{-4ix}) - \cdots\right).$$

The function $g(x)$ has a very similar development. For both functions the series of absolute values are uniformly convergent. Hence the expansion (5) converges to $f(x)$; likewise for $g(x)$. The two series may be multiplied and terms rearranged and grouped in any convenient fashion; their product is

$$0 = f(x)\cdot g(x) = \sum_{-\infty}^{\infty} a_r e^{irx}.$$

By the Heine-Cantor theorem, $a_r = 0$, $-\infty < r < \infty$. As a matter of fact a very elementary reasoning involving term-by-term integration gives $a_r = 0$.

In the series (5) replace exp (ix) by V. The resulting series of operators is absolutely convergent since $|V^n| \leqq K$. Call the transformation which the series defines A. That is,

$$(6) \quad A = \frac{1}{4i}(V - V^{-1}) + \frac{1}{\pi}\left(I - \frac{1}{1\cdot 3}(V^2 + V^{-2}) - \frac{1}{3\cdot 5}(V^4 + V^{-4}) - \cdots\right).$$

Similarly, corresponding to the expansion of $g(x)$, write

$$(7) \quad B = \frac{1}{4i}(V - V^{-1}) - \frac{1}{\pi}\left(I - \frac{1}{1\cdot 3}(V^2 + V^{-2}) - \frac{1}{3\cdot 5}(V^4 + V^{-4}) - \cdots\right).$$

The transformation $A\cdot B$ is obtained by multiplying the above series. Since the operations here parallel closely those for evaluating $f(x)\cdot g(x)$, it follows that

$$(8) \qquad A\cdot B = B\cdot A = 0.$$

Two theorems will now be introduced which will be vital to the following discussion. As their proof has been given by the author elsewhere, only their statement will be reproduced here(7).

THEOREM 3. *Let $\mathfrak{B}$ be a reflexive vector space. Let T be a bounded linear transformation in $\mathfrak{B}$. Let $\mathfrak{M}$ denote the closed linear manifold of elements f for which $Tf = 0$; let $\mathfrak{N}$ denote the closed linear manifold spanned by the elements Tf, $f \in \mathfrak{B}$.*

(7) Lorch, *Means of iterated transformations in reflexive vector spaces*, Bulletin of the American Mathematical Society, vol. 45 (1939), pp. 945–947.

Let $\bar{T}$ be the adjoint of T, and let $(\mathfrak{M})$ and $(\mathfrak{N})$ be the manifolds associated with $\bar{T}$ in the manner just indicated. Then $(\mathfrak{M})=\mathfrak{N}^{\perp}$, $(\mathfrak{N})=\mathfrak{M}^{\perp}$, $\mathfrak{M}=(\mathfrak{N})^{\perp}$, $\mathfrak{N}=(\mathfrak{M})^{\perp}$.

THEOREM 4. *Let $\mathfrak{B}$ be a reflexive vector space. Let W be a bounded linear transformation in $\mathfrak{B}$ such that $|W^n|\leqq K$, $n=0, 1, 2, \cdots$. Then the sequence $\{T_n\}$ where $T_n=(1/n)\sum_{i=0}^{n-1}W^i$ converges strongly to a limiting transformation P which is a projection. The relation $Pf=f$ holds precisely for those elements f for which $Wf=f$. The relation $Pf=0$ holds precisely for the elements of the closed linear manifold spanned by elements of the form $Wg-g$, g arbitrary in $\mathfrak{B}$. The bound of P satisfies the inequality $|P|\leqq K$.*

It is to be noted that if $\mathfrak{M}$ is a closed linear manifold (for short, c.l.m.) in $\mathfrak{B}$, then $\mathfrak{M}^{\perp}$ denotes the c.l.m. of elements of $(\mathfrak{B})$, each of which is orthogonal to every element of $\mathfrak{M}$. Also to be noted is the fact that the transformation W of Theorem 4 is not required to be bicontinuous.

Slight changes in the hypotheses of Theorem 4 lead to a result which is useful later. The result is valid in any Banach space.

THEOREM 4'. *Let W be a transformation in an arbitrary Banach space $\mathfrak{B}$ such that the norm of W^n satisfies the inequality $|W^n|\leqq a^nK$, $n=1, 2, 3, \cdots$, with $a<1$, $K\geqq 0$. Then $W-I$ possesses a bounded inverse $(W-I)^{-1}$.*

A simple proof, based on the Neumann development, may be given. First, if $f\in\mathfrak{B}$, there exists an element g such that $f=g-Wg$. For let

$$f+Wf+W^2f+\cdots=g.$$

The convergence of the series is assured since, for a given $\epsilon>0$,

$$\begin{aligned}\|W^nf+W^{n+1}f+\cdots+W^mf\| &\leqq \|W^nf\|+\cdots+\|W^mf\|\\ &\leqq K\|f\|(a^n+\cdots+a^m)<\epsilon\end{aligned}$$

providing n is sufficiently large. Thus

$$Wf+W^2f+\cdots=Wg$$

and $f=g-Wg$. Hence the range of $W-I$ is $\mathfrak{B}$. Further, if $(W-I)h=0$, then $Wh=h$, $W^nh=h$, and $h=0$ since $|W^n|\leqq a^nK$. Thus $W-I$ maps $\mathfrak{B}$ into itself in a one-to-one way and is continuous (in one direction). Hence $W-I$ possesses a bounded inverse[8].

Now let $\mathfrak{B}$ be reflexive once more and V be w.a.p. Applying Theorem 4 to the transformation V^2 (which is uniformly bounded), it is seen that the space $\mathfrak{B}$ is the "sum" of two disjoint[9] manifolds, $\mathfrak{B}=\mathfrak{M}\dot{+}\mathfrak{N}$ with $V^2f=f$ for

[8] Banach, p. 41.

[9] Disjointness is treated in the author's paper *On a calculus of operators* $\cdots$, p. 220. For short, one may say that two c.l.m.'s $\mathfrak{M}$ and $\mathfrak{N}$ are disjoint if they have in common only the element 0 and if the totality of elements $f+g$ with $f\in\mathfrak{M}$ and $g\in\mathfrak{N}$ is a c.l.m.

every $f \varepsilon \mathfrak{M}$ and with $\mathfrak{N}$ spanned by the elements $(V^2-I)g$, g arbitrary, or, what amounts to the same, by the elements $(V-V^{-1})g$. The reader is reminded that the c.l.m. spanned by a set of elements is the smallest c.l.m. containing the set. For any $f \varepsilon \mathfrak{M}$ write

$$f = f_1 + f_2, \qquad f_1 = \frac{f+Vf}{2}, \qquad f_2 = \frac{f-Vf}{2};$$

note that $Vf_1=f_1$; $Vf_2=-f_2$. Thus if $\mathfrak{M}'$ denotes the c.l.m. of elements f_1 and $\mathfrak{M}''$ denotes the c.l.m. of elements f_2, $\mathfrak{M}=\mathfrak{M}'\dot{+}\mathfrak{M}''$, the latter manifolds being disjoint. Thus if f is now arbitrary in $\mathfrak{B}$, the decomposition $f=g_1+g_2+h$ with $g_1 \varepsilon \mathfrak{M}'$, $g_2 \varepsilon \mathfrak{M}''$, $h \varepsilon \mathfrak{N}$ is possible and unique. This fact will be of the greatest utility subsequently.

The hypotheses of the following theorem may always be fulfilled for any w.a.p. V by considering the transformation in the c.l.m. $\mathfrak{N}$. The theorem gives the fundamental partition of $\mathfrak{B}$ whose subsequent refinements yield the key to the structure of V.

THEOREM 5. *Let V be a weakly almost-periodic transformation in a reflexive space $\mathfrak{B}$. Let the equation $V^2f=f$ have the unique solution $f=0$. Let A and B be the linear transformations introduced in* (6) *and* (7) *respectively. Let $\mathfrak{Q}$ be the closed linear manifold of elements f for which $Af=0$; let $\mathfrak{R}$ be the closed linear manifold of elements f for which $Bf=0$. Then $\mathfrak{Q}$ and $\mathfrak{R}$ have only the element* 0 *in common; together they span $\mathfrak{B}$. The manifold pair $\{\mathfrak{Q}, \mathfrak{R}\}$ reduces V in the sense that if $g \varepsilon \mathfrak{Q}$, $Vg \varepsilon \mathfrak{Q}$; if $h \varepsilon \mathfrak{R}$, $Vh \varepsilon \mathfrak{R}$.*

Since $A\cdot B=B\cdot A=0$ by (8), the c.l.m. $\mathfrak{Q}'$ spanned by the elements Bf, $f \varepsilon \mathfrak{B}$ is included in $\mathfrak{Q}$. Similarly, the c.l.m. $\mathfrak{R}'$ spanned by the elements Af is included in $\mathfrak{R}$.

Applying Theorem 4, we see that since $(V^2-I)f=0$ implies $f=0$, the space $\mathfrak{B}$ is spanned by the elements of form $(V-V^{-1})f$, $f \varepsilon \mathfrak{B}$. Hence the elements $(A+B)f$, $f \varepsilon \mathfrak{B}$. Since $Af \varepsilon \mathfrak{R}$, $Bf \varepsilon \mathfrak{Q}$, the manifolds $\mathfrak{Q}$ and $\mathfrak{R}$ span $\mathfrak{B}$.

If $f \varepsilon \mathfrak{Q}$, $f \varepsilon \mathfrak{R}$, then $Af=Bf=0$ or $(A+B)f=0$. Hence $f=0$; thus $\mathfrak{Q}$ and $\mathfrak{R}$ have only the element 0 in common.

Since $AV=VA$ and $BV=VB$ (see (6) and (7)), then $Ag=0$ implies $AVg=0$; $Bh=0$ implies $BVh=0$. Thus the manifold pair $\{\mathfrak{Q}, \mathfrak{R}\}$ reduces V in the sense explained above.

It may quickly be seen that the manifolds $\mathfrak{Q}'$ and $\mathfrak{R}'$ have all the properties which the theorem ascribes to $\mathfrak{Q}$ and $\mathfrak{R}$. In case $\mathfrak{Q}'$ and $\mathfrak{R}'$ are disjoint, $\mathfrak{Q}=\mathfrak{Q}'$ and $\mathfrak{R}=\mathfrak{R}'$. For if $\mathfrak{Q}'$ and $\mathfrak{R}'$ are disjoint, for an arbitrary $f \varepsilon \mathfrak{B}$ one may find a $g \varepsilon \mathfrak{Q}'$ and an $h \varepsilon \mathfrak{R}'$ such that $f=g+h$. If now in particular $f \varepsilon \mathfrak{Q}$, then $f-g=h \varepsilon \mathfrak{Q}$ as well as $h \varepsilon \mathfrak{R}'$. Since $\mathfrak{R} \supset \mathfrak{R}'$, $h=0$ and $f \varepsilon \mathfrak{Q}'$. Thus $\mathfrak{Q}=\mathfrak{Q}'$; similarly $\mathfrak{R}=\mathfrak{R}'$[10].

[10] If the c.l.m.'s $\mathfrak{Q}$ and $\mathfrak{R}$ are not disjoint, it is easy to construct a c.l.m. $\mathfrak{S}$ of which $\mathfrak{Q}$ is a proper subset and such that $\mathfrak{R}$ and $\mathfrak{S}$ have only the element 0 in common. For let f be any

In the special case in which V is a unitary transformation in a Hilbert space, $\mathfrak{Q}$ and $\mathfrak{N}$ may be shown to be orthogonal in the following way. Using the fact that $\overline{V}=V^{-1}$, it may be seen by direct computation in (6) that $A=\overline{A}$ (A is self-adjoint). Thus the c.l.m. spanned by the elements $\overline{A}f$ is $\mathfrak{N}'$. By Theorem 3, this c.l.m. is the orthogonal complement of $\mathfrak{Q}$. Thus $\mathfrak{Q}'$ and $\mathfrak{N}'$ are orthogonal to each other, hence disjoint. By the previous paragraph $\mathfrak{Q}=\mathfrak{Q}'$, $\mathfrak{N}=\mathfrak{N}'$.

Attempts to determine whether the manifolds $\mathfrak{Q}$ and $\mathfrak{N}$ are disjoint in general have not been successful. The following theorem gives some criteria which may be useful.

THEOREM 6. *Let V be a weakly almost-periodic transformation satisfying the conditions given in Theorem 5. The closed linear manifolds $\mathfrak{Q}$ and $\mathfrak{N}$ there defined are disjoint if either of the following conditions is satisfied:*

(a) *The transformation $A+B$ possesses a bounded inverse $(A+B)^{-1}$.*

(b) *The transformation $I-A$ (or $I-B$) is uniformly bounded, $|(I-A)^n| \leqq L$, $n=1, 2, \cdots$.*

If $(A+B)^{-1}$ exists and is bounded, then for any $f \varepsilon \mathfrak{B}$,

$$f=(A+B)(A+B)^{-1}f=A(A+B)^{-1}f+B(A+B)^{-1}f$$

which gives the resolution of f into two elements, the first in $\mathfrak{N}$, the second in $\mathfrak{Q}$. Since $\mathfrak{Q}$ and $\mathfrak{N}$ have only the element 0 in common, this resolution is unique. Thus $\mathfrak{Q}$ and $\mathfrak{N}$ are disjoint.

Suppose that the transformation $I-A$ is uniformly bounded, $|(I-A)^n| \leqq L$, $n=1, 2, \cdots$. Then by Theorem 4, the two following manifolds are disjoint: the c.l.m. of elements f for which $(I-A)f=f$, that is, the c.l.m. $\mathfrak{Q}$; and the c.l.m. spanned by the elements $[(I-A)-I]f=-Af$, that is, the c.l.m. $\mathfrak{N}'$. Thus $\mathfrak{Q}'$ and $\mathfrak{N}'$ are disjoint and $\mathfrak{Q}=\mathfrak{Q}'$ $\mathfrak{N}=\mathfrak{N}'$.

In this connection, it may be pointed out that the uniform boundedness of the transformations $A+B$, A, and B may be proved easily. Since $A+B=(V-V^{-1})/2i$, it may be seen by direct multiplication that $|(A+B)^n| \leqq K$ where K is a uniform bound for V^n. Since $A^n=A(A+B)^{n-1}$ and $B^n=B(A+B)^{n-1}$, A and B are uniformly bounded. Similarly the uniform boundedness of the transformation $I-(A+B)^2$ may be demonstrated. Also demonstrable is the uniform boundedness on the c.l.m. $\mathfrak{Q}'$ or on the c.l.m. $\mathfrak{N}'$ of the transformation $I-A^2$.

If the manifolds $\mathfrak{Q}$ and $\mathfrak{N}$ are disjoint and if certain even more stringent conditions involving uniformity of bound are satisfied, it is possible to con-

element which cannot be represented in the form $g+h$ with $g \varepsilon \mathfrak{Q}$, $h \varepsilon \mathfrak{N}$. Then the totality $\mathfrak{S}$ of elements of the form $\alpha f+g$, $g \varepsilon \mathfrak{Q}$ is a c.l.m. having the required properties. That $\mathfrak{S}$ is closed may be seen as follows: Let $\{\alpha_n f+g_n\}$ be a convergent sequence of elements in $\mathfrak{S}$. Let F be a linear functional which is orthogonal to $\mathfrak{Q}$ and for which $Ff=1$. Then $F(\alpha_n f+g_n)=\alpha_n \rightarrow \alpha$. Hence $g_n \rightarrow g \varepsilon \mathfrak{Q}$ and $\alpha_n f+g_n \rightarrow \alpha f+g \varepsilon \mathfrak{S}$.

struct for V a resolution of the identity[11]. The present situation, however, leads to a structure much more general it seems than that of a resolution of the identity. It is described in the next theorem. This theorem completes the analysis or decomposition of V; the reconstruction of V from the manifold pairs which reduce it will be undertaken subsequently.

A useful lemma will first be established. Let λ be so chosen that $-\pi<\lambda\leqq\pi$. The transformation $V_\lambda=\exp(-i\lambda)V$ is w.a.p. In what follows, A_λ and B_λ will denote the transformations which are constructed from V_λ in the manner in which $A=A_0$ and $B=B_0$ are constructed from $V=V_0$.

LEMMA. *Let $\lambda_1, \lambda_2, \cdots, \lambda_n$ be any distinct real numbers with $-\pi<\lambda_j\leqq 0$. Then for any w.a.p. transformation V and for every $f \in \mathfrak{B}$, the following decomposition is possible and unique*:

$$f=\sum_{j=1}^{n}(g_{\lambda_j}+g'_{\lambda_j})+h \tag{9}$$

with

$$Vg_{\lambda_j}=e^{i\lambda_j}g_{\lambda_j}, \qquad Vg'_{\lambda_j}=-e^{i\lambda_j}g'_{\lambda_j},$$

and with h in the closed linear manifold spanned by the elements of the form

$$Xk=(A_{\lambda_1}+B_{\lambda_1})(A_{\lambda_2}+B_{\lambda_2})\cdots(A_{\lambda_n}+B_{\lambda_n})k,$$

k arbitrary in $\mathfrak{B}$.

Let the projection which Theorem 4 defines for V_λ be denoted by P_λ; then $V_\lambda P_\lambda f=P_\lambda f$ or $VP_\lambda f=\exp(i\lambda)P_\lambda f$. Let the projection which is defined for $-V_\lambda=V_{\lambda+\pi}$ be denoted by P'_λ; then $VP'_\lambda f=-\exp(i\lambda)P'_\lambda f$. Note that the projections of the collection P_λ and P'_λ for all λ, $-\pi<\lambda\leqq 0$, are commutative. Indeed, the product of two distinct projections is 0. Hence

$$\begin{aligned} I &= \prod_{j=1}^{n}[P_{\lambda_j}+(I-P_{\lambda_j})][P'_{\lambda_j}+(I-P'_{\lambda_j})] \\ &= \sum_{j=1}^{n}(P_{\lambda_j}+P'_{\lambda_j})+\prod_{j=1}^{n}(I-P_{\lambda_j})(I-P'_{\lambda_j}). \end{aligned}$$

This gives a resolution for $f \in \mathfrak{B}$ of the type indicated in the lemma except for the fact that the announced properties of h have not yet been established. To prove that the resolution is unique let $\sum_{j=1}^{n}(g_{\lambda_j}+g'_{\lambda_j})+h=0$; operating with P_{λ_r}, one obtains $P_{\lambda_r}g_{\lambda_r}=g_{\lambda_r}=0$ for $r=1, 2, \cdots, n$. Similarly $g'_{\lambda_r}=0$, $r=1, \cdots, n$, hence finally $h=0$. The resolution is thus unique.

It remains to be shown that h has the character specified in the lemma. Since h is of the form $(I-P_{\lambda_j})k$, by Theorem 4 h is in the c.l.m. spanned by the elements of the form $(V_{\lambda_j}-I)l$; similarly, it is in the c.l.m. spanned by the ele-

(11) Lorch, *On a calculus* $\cdots$, p. 226, Definition 4.

ments of the form $(-V_{\lambda_j}-I)l$. Now the intersection of these two manifolds includes the c.l.m. spanned by the elements of the form $(A_{\lambda_j}+B_{\lambda_j})l$. If $Y=\prod_{j=1}^{n}(I-P_{\lambda_j})(I-P'_{\lambda_j})$, then Y is a projection and the c.l.m. of elements Yl is precisely the intersection of the c.l.m.'s spanned by the elements $(V_{\lambda_j}-I)l$ and $(-V_{\lambda_j}-I)l$, $j=1, \cdots, n$. Hence the c.l.m. of all elements Yl includes the c.l.m. spanned by the elements Xl.

It will be shown that the manifold spanned by the elements Xl is identical with the manifold of the elements Yl. To do this it is necessary and sufficient to establish that the collection of elements $\sum_{j=1}^{n}(g_{\lambda_j}+g'_{\lambda_j})+Xl$ is dense in $\mathfrak{B}$. If $f \varepsilon \mathfrak{B}$, then f may be approximated in the norm by elements $g_{\lambda_1}+g'_{\lambda_1}+(A_{\lambda_1}+B_{\lambda_1})h_1$ (see discussion preceding Theorem 5). Now h_1 may be approximated by elements $g_{\lambda_2}+g'_{\lambda_2}+(A_{\lambda_2}+B_{\lambda_2})h_2$. Since $A_{\lambda_1}g_{\lambda_2}=f(\lambda_2-\lambda_1)g_{\lambda_2}$ (that is, $A_{\lambda_1}g_{\lambda_2}=\sin(\lambda_2-\lambda_1)g_{\lambda_2}$ or 0), and $A_{\lambda_1}g'_{\lambda_2}=f(\pi+\lambda_2-\lambda_1)g'_{\lambda_2}$, and since $B_{\lambda_1}g_{\lambda_2}=g(\lambda_2-\lambda_1)g_{\lambda_2}$, $B_{\lambda_1}g'_{\lambda_2}=g(\pi+\lambda_2-\lambda_1)g'_{\lambda_2}$, f may be approximated by elements of the form $g_{\lambda_1}+g'_{\lambda_1}+\rho g_{\lambda_2}+\sigma g'_{\lambda_2}+(A_{\lambda_1}+B_{\lambda_1})(A_{\lambda_2}+B_{\lambda_2})h_2$, where ρ and σ are some constants. Now let h_2 be approximated by an element $g_{\lambda_3}+g'_{\lambda_3}+(A_{\lambda_3}+B_{\lambda_3})h_3$, etc. Proceeding in this way, it can be seen that any element f can be approximated by elements of the form $\sum(g_{\lambda_j}+g'_{\lambda_j})+Xl$.

Note that if $\mathfrak{M}$ is a c.l.m. which has the property that it contains Vf whenever it contains f, then along with f, $\mathfrak{M}$ contains g_{λ_j}, g'_{λ_j}, $j=1, \cdots, n$, and h. For

$$g_{\lambda_j} = P_{\lambda_j}f = \lim_{n\to\infty} \frac{1}{n} \sum_{s=0}^{n-1} V_{\lambda_j}^{s} f;$$

hence $g_{\lambda_j} \varepsilon \mathfrak{M}$. Similarly the g'_{λ_j} and thus h are in $\mathfrak{M}$.

If in (9) one operates upon f with V, one obtains the decomposition of Vf, that is, $Vf=\sum_{j=1}^{n}(Vg_{\lambda_j}+Vg'_{\lambda_j})+Vh$ is the decomposition (9) for the element Vf. This is because the projections P_{λ_j}, P'_{λ_j} are commutative with V. What is here stated for V applies equally well to any function of V, say , A, B, etc.

THEOREM 7. *Let V be a weakly almost-periodic transformation in a reflexive vector space $\mathfrak{B}$. Then for every real number λ, $-\infty<\lambda<\infty$, there exists a pair of closed linear manifolds, $\{\mathfrak{E}_\lambda, \mathfrak{F}_\lambda\}$, possessing the following properties:*

(a) *The pair $\{\mathfrak{E}_\lambda, \mathfrak{F}_\lambda\}$ reduces V in the sense of Theorem* 5.

(b) *The manifolds $\mathfrak{E}_\lambda$ and $\mathfrak{F}_\lambda$ have in common only the element* 0; *together, $\mathfrak{E}_\lambda$ and $\mathfrak{F}_\lambda$ span $\mathfrak{B}$.*

(c) *$\mathfrak{E}_\lambda \supset \mathfrak{E}_\mu$ for $\lambda>\mu$; $\mathfrak{F}_\lambda \subset \mathfrak{F}_\mu$ for $\lambda>\mu$.*

(d) $\mathfrak{E}_{-\pi}=0$, $\mathfrak{F}_{-\pi}=\mathfrak{B}$; $\mathfrak{E}_{\pi}=\mathfrak{B}$, $\mathfrak{F}_{\pi}=0$.

Preliminary to constructing the $\mathfrak{E}_\lambda$ and $\mathfrak{F}_\lambda$ certain other manifolds $\mathfrak{M}_\lambda$ and $\mathfrak{N}_\lambda$, $-\pi<\lambda\leqq\pi$, will be introduced. For $-\pi<\lambda\leqq 0$, apply the lemma just established to the case $n=1$, $\lambda_1=\lambda$; thus $f=g_\lambda+g'_\lambda+h$ with $Vg_\lambda=\exp(i\lambda)g_\lambda$, $Vg'_\lambda=-\exp(i\lambda)g'_\lambda$, and with h in the c.l.m. spanned by the elements $(A_\lambda+B_\lambda)l$. Let $\mathfrak{M}_\lambda$ be the collection of all elements f such that $A_\lambda f=0$ and

such that $g_\lambda' = 0$. Let $\mathfrak{N}_\lambda$ be the collection of all elements f such that $B_\lambda f = 0$ and such that $g_\lambda = 0$. In the first case, from $A_\lambda f = 0$ follows $A_\lambda h = 0$. Thus $\mathfrak{M}_\lambda$ is the c.l.m. which is the direct sum of the following two disjoint manifolds: The c.l.m. of all elements g_λ for which $Vg_\lambda = \exp(i\lambda)g_\lambda$; the c.l.m. of all elements h for which $A_\lambda h = 0$. Similarly, $\mathfrak{N}_\lambda$ is the c.l.m. which is the direct sum of the two disjoint manifolds: The c.l.m. of all elements g_λ' for which $Vg_\lambda' = -\exp(i\lambda)g_\lambda'$; the c.l.m. of all elements h for which $B_\lambda h = 0$.

For $0 < \lambda \leqq \pi$, let $\mathfrak{M}_\lambda$ and $\mathfrak{N}_\lambda$ be defined by $\mathfrak{M}_\lambda = \mathfrak{N}_{\lambda-\pi}$, $\mathfrak{N}_\lambda = \mathfrak{M}_{\lambda-\pi}$. Since $A_\lambda = -B_{\lambda-\pi}$, $B_\lambda = -A_{\lambda-\pi}$, $\mathfrak{M}_\lambda$ is precisely the c.l.m. which is the direct sum of two c.l.m.'s: the c.l.m. of the elements g for which $Vg = \exp(i\lambda)g$; and the c.l.m. of the elements h for which $A_\lambda h = 0$. A similar statement may be made for $\mathfrak{N}_\lambda$.

The pair $\{\mathfrak{M}_\lambda, \mathfrak{N}_\lambda\}$ reduces V. Assume $-\pi < \lambda \leqq 0$. If $f \,\varepsilon\, \mathfrak{M}_\lambda$, $f = g_\lambda + h$, $Vf = Vg_\lambda + Vh$, $A_\lambda Vf = VA_\lambda f = 0$, hence $Vf \,\varepsilon\, \mathfrak{M}_\lambda$. Similarly for $\mathfrak{N}_\lambda$.

The manifolds $\mathfrak{M}_\lambda$ and $\mathfrak{N}_\lambda$ have only the element 0 in common. Let $f \,\varepsilon\, \mathfrak{M}_\lambda$, $f \,\varepsilon\, \mathfrak{N}_\lambda$, $f = g_\lambda + h$, $f = g_\lambda' + k$. Then $0 = g_\lambda - g_\lambda' + h - k$, $g_\lambda = g_\lambda' = 0$, $f = h = k$ and $A_\lambda h = B_\lambda h = (A_\lambda + B_\lambda)h = 0$; thus $h = 0$.

The manifolds $\mathfrak{M}_\lambda$ and $\mathfrak{N}_\lambda$ span $\mathfrak{B}$. For the elements $g_\lambda + g_\lambda' + (A_\lambda + B_\lambda)k$ lie dense in $\mathfrak{B}$; all elements g_λ and $B_\lambda k$ lie in $\mathfrak{M}_\lambda$; the elements g_λ' and $A_\lambda k$ lie in $\mathfrak{N}_\lambda$.

The manifolds $\mathfrak{E}_\lambda$, $\mathfrak{F}_\lambda$, $-\infty < \lambda < \infty$, will be defined. For $-\infty < \lambda \leqq -\pi$, $\mathfrak{E}_\lambda = 0$ and $\mathfrak{F}_\lambda = \mathfrak{B}$. For $\pi \leqq \lambda < \infty$, $\mathfrak{E}_\lambda = \mathfrak{B}$ and $\mathfrak{F}_\lambda = 0$. In discussing the theorem, these values of λ will receive no attention.

Let $-\pi < \lambda < 0$. Consider the partition for $f \,\varepsilon\, \mathfrak{B}$ of the lemma for $n = 2$, $\lambda_1 = 0$, $\lambda_2 = \lambda$:

$$f = g_0 + g_0' + g_\lambda + g_\lambda' + h.$$

The c.l.m. $\mathfrak{E}_\lambda$ is defined to be the collection of elements f for which $Af = A_\lambda f = 0$, and such that $g_0 = g_0' = g_\lambda' = 0$. The c.l.m. $\mathfrak{F}_\lambda$ is defined to be the collection of elements f for which $BB_\lambda f = 0$ and such that $g_\lambda = 0$.

If $0 < \lambda < \pi$, write $\mu = \lambda - \pi$ and consider the partition for f of the lemma corresponding to $n = 2$, $\lambda_1 = 0$, $\lambda_2 = \mu$. Then $\mathfrak{E}_\lambda$ is defined to be the c.lm. of elements f for which $AB_\mu f = -AA_\lambda f = 0$ and such that $g_0' = 0$. Also, $\mathfrak{F}_\lambda$ is defined to be c.l.m. of elements f for which $Bf = 0$, $A_\mu f = 0$ (note that $A_\mu = -B_\lambda$), and such that $g_0 = g_\mu = g_\mu' = 0$.

For $\lambda = 0$, $\mathfrak{E}_0 = \mathfrak{M}_0$ and $\mathfrak{F}_0 = \mathfrak{N}_0$ by definition.

It may be noted that for $-\pi < \lambda \leqq 0$, $\mathfrak{E}_\lambda$ is precisely the intersection of $\mathfrak{M}_0$ and $\mathfrak{M}_\lambda$; and $\mathfrak{F}_\lambda$ contains $\mathfrak{N}_0$ and $\mathfrak{N}_\lambda$.

Toward (a). Let $-\pi < \lambda < 0$. As stated above, the partition (9) for Vf is obtained from that of f by operating upon (9) with V. Since $AV = VA$, and $A_\lambda V = VA_\lambda$, whenever $f \,\varepsilon\, \mathfrak{E}_\lambda$, then $Vf \,\varepsilon\, \mathfrak{E}_\lambda$. Similarly since $VBB_\lambda = BB_\lambda V$, if $f \,\varepsilon\, \mathfrak{F}_\lambda$, then $Vf \,\varepsilon\, \mathfrak{F}_\lambda$.

Toward (b). Suppose $f \varepsilon \mathfrak{E}_\lambda$, $f=g_\lambda+h$; if $f \varepsilon \mathfrak{F}_\lambda$, then $f=g_0+g_0'+g_\lambda'+k$. Thus $0=g_0+g_0'-g_\lambda+g_\lambda'+k-h$ and $g_0=g_0'=g_\lambda=g_\lambda'=0$, $f=k=h$. Further, since $Ah=A_\lambda h=BB_\lambda h=0$, $(A+B)(A_\lambda+B_\lambda)h=0$ and $h=0$. Thus $\mathfrak{E}_\lambda$ and $\mathfrak{F}_\lambda$ have only the element 0 in common. To prove that $\mathfrak{E}_\lambda$ and $\mathfrak{F}_\lambda$ span $\mathfrak{B}$, note that $g_\lambda \varepsilon \mathfrak{E}_\lambda$, g_0, g_0', $g_\lambda' \varepsilon \mathfrak{F}_\lambda$, and also that h may be approximated by elements of the form $(A+B)(A_\lambda+B_\lambda)k$ with $BB_\lambda k$ in $\mathfrak{E}_\lambda$ and with $AA_\lambda k$, $BA_\lambda k$, $AB_\lambda k$ in $\mathfrak{F}_\lambda$.

Toward (c). Let $-\pi<\lambda<\mu<0$. Let $f \varepsilon \mathfrak{E}_\lambda$; it will be shown that $f \varepsilon \mathfrak{E}_\mu$. Note that $BB_\lambda A_\mu=0$ identically in $\mathfrak{B}$. This may be seen as follows: The transformation B is constructed with the help of the function $\mathfrak{g}(x)$—see (5), (6), and (7). Similarly, the transformations A_μ and B_λ are constructed with the help of the functions $\mathfrak{f}(x-\mu)$ and $\mathfrak{g}(x-\lambda)$ respectively. Now $\mathfrak{g}(x)\cdot\mathfrak{g}(x-\lambda)\cdot\mathfrak{f}(x-\mu)=0$ identically in x. This product may be evaluated by multiplying together three Fourier series, yielding a product series having zero coefficients. Precisely the same series operations occur in constructing the transformation $BB_\lambda A_\mu$ except that $\exp(ix)$ is replaced by V. Thus $BB_\lambda A_\mu=0$. Since $f \varepsilon \mathfrak{E}_\lambda$, $Af=A_\lambda f=0$ and

$$(A + B)(A_\lambda + B_\lambda)A_\mu f = BB_\lambda A_\mu f = 0.$$

Let $f=g_\lambda+h$ be the decomposition (9) of f for $n=2$, $\lambda_1=0$, $\lambda_2=\lambda$. Then $A_\mu f=A_\mu g_\lambda+A_\mu h=A_\mu h$ since $A_\mu g_\lambda=\mathfrak{f}(\lambda-\mu)g_\lambda=0$. Also $(A+B)(A_\lambda+B_\lambda)A_\mu h=0$ implies $A_\mu h=0$, that is, $A_\mu f=0$. Let $f=g_0+g_0'+g_\mu+g_\mu'+k$ be the decomposition (9) of f for $n=2$, $\lambda_1=0$, $\lambda_2=\mu$. Since $f \varepsilon \mathfrak{E}_\lambda$, $g_0=g_0'=0$. Further since $Af=0$, $Ag_\mu'=\sin(\mu+\pi)g_\mu'=0$, and $g_\mu'=0$. Therefore $f \varepsilon \mathfrak{E}_\mu$ and $\mathfrak{E}_\lambda \subset \mathfrak{E}_\mu$.

It will now be proved that $\mathfrak{F}_\mu \subset \mathfrak{F}_\lambda$ (for $-\pi<\lambda<\mu<0$). Let $f \varepsilon \mathfrak{F}_\mu$; then $BB_\mu f=0$. Also,

$$(A_\mu + B_\mu)BB_\lambda f = B_\mu BB_\lambda f = 0$$

since $A_\mu BB_\lambda=0$. Let $f=g_0+g_0'+g_\mu+g_\mu'+h$ be the decomposition (9) for $n=2$, $\lambda_1=0$, $\lambda_2=\mu$. Since $f \varepsilon \mathfrak{F}_\mu$, $g_\mu=0$. Thus $BB_\lambda f=BB_\lambda g_\mu'+BB_\lambda h=BB_\lambda h$ (since $Bg_\mu'=\mathfrak{g}(\mu+\pi)g_\mu'=0$) is the decomposition (9) of $BB_\lambda f$. Applying the operator $(A_\mu+B_\mu)$, one sees that $BB_\lambda f=0$. Now let $f=g_0+g_0'+g_\lambda+g_\lambda'+h$ be the partition of f relative to $n=2$, $\lambda_1=0$, $\lambda_2=\lambda$. Since $BB_\mu f=0$, $BB_\mu g_\lambda=0$, that is, $\mathfrak{g}(\lambda)\cdot\mathfrak{g}(\lambda-\mu)g_\lambda=0$, or $g_\lambda=0$. Hence $f \varepsilon \mathfrak{F}_\lambda$, and $\mathfrak{F}_\mu \subset \mathfrak{F}_\lambda$.

The method of proof of the properties (b) and (c) for other values (and pairs of values) of λ, for instance for $0<\lambda<\pi$, is now clear and will not be set down here. The proof of (d) lies in the definition of $\mathfrak{E}_\lambda$ and $\mathfrak{F}_\lambda$ for $\lambda=-\pi, \pi$.

IV. The integral representation of V

The aim of this chapter is to establish an integral representation (2) for the w.a.p. transformation V. The precise interpretation of (2) in the present case will be given in the next theorem. Preparatory to this theorem, a few auxiliary devices will be investigated.

Let $\epsilon_2>0$ be any positive number. Suppose that for $f \epsilon \mathfrak{B}$, $(V-I)^2f=\eta$ with $\|\eta\| \leqq \epsilon_1\|f\|$ for some $\epsilon_1>0$. Then

$$(10) \qquad \begin{aligned} (V^2 - V)f &= (V - I)f + \eta, \\ (V^3 - V^2)f &= (V - I)f + V\eta + \eta, \\ &\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots, \\ (V^{n+1} - V^n)f &= (V - I)f + V^{n-1}\eta + \cdots + V\eta + \eta. \end{aligned}$$

Thus

$$n(V - I)f = (V^{n+1} - V)f - n\eta - (n - 1)V\eta - \cdots - V^{n-1}\eta.$$

or

$$\|(V - I)f\| \leqq \frac{2}{n} K\|f\| + \frac{n + 1}{2} K\epsilon_1\|f\|.$$

If now n is chosen so large that $2K/n \leqq \epsilon_2/2$, and if it happens for that n that $(n+1)K\epsilon_1 \leqq \epsilon_2$, then $\|(V-I)f\| \leqq \epsilon_2\|f\|$.

Suppose now that for some $f \epsilon \mathfrak{B}$, $(V^2-I)f=\eta$ with $\|\eta\| \leqq \epsilon\|f\|$ for some $\epsilon>0$. Then writing $f=f_1+f_2$ with

$$(11) \qquad f_1 = \frac{f + Vf}{2}, \qquad f_2 = \frac{f - Vf}{2},$$

it is seen that $(V-I)f_1=\eta/2$, $(V+I)f_2=-\eta/2$. Thus it follows if ϵ is "small" that f is the sum of two elements for which V is "almost" the identity transformation or its negative. The following theorem indicates precisely in what sense the transformation V may be reconstructed from $\{\mathfrak{E}_\lambda, \mathfrak{F}_\lambda\}$.

THEOREM 8. *Let a real number λ and an $\epsilon>0$ be given. Then it is possible to find a closed linear manifold $\mathfrak{A}_\lambda$ such that:*

(a) *The manifold $\mathfrak{A}_\lambda$ lies in $\mathfrak{E}_{\lambda+\epsilon}$ and $\mathfrak{F}_{\lambda-\epsilon}$.*

(b) *If $f \epsilon \mathfrak{A}_\lambda$, the inequality*

$$(12) \qquad \|(V - e^{i\lambda}I)f\| \leqq \epsilon\|f\|$$

holds.

Furthermore there exists a finite collection of values λ, viz., $\lambda_1, \lambda_2, \cdots, \lambda_n$, such that the space $\mathfrak{B}$ is spanned by the manifolds $\mathfrak{A}_{\lambda_i}$, $i=1, 2, \cdots, n$.

Consider the transformation $A_{\lambda-\delta}\cdot B_{\lambda+\delta}$, with λ arbitrary but fixed, $\delta>0$ arbitrary. Its bound is a continuous function of δ which is zero for $\delta=0$. Writing $\lambda-\delta=\alpha$, $\lambda+\delta=\beta$,

$$A_\alpha B_\beta \cdot A_\alpha B_\beta = (A_\alpha + B_\alpha)(A_\beta + B_\beta)A_\alpha B_\beta.$$

By choosing δ sufficiently small, the norm of

$$(A_\lambda + B_\lambda)^2 - (A_\alpha + B_\alpha)(A_\beta + B_\beta)$$

may be made less than $\epsilon_1/2$ where $\epsilon_1>0$ is arbitrary. This choice of δ is independent of λ. If at the same time δ is so small that $|A_\alpha B_\beta| \leqq \epsilon_1/2$, then

$$\|(A_\lambda + B_\beta)^2 A_\alpha B_\lambda f\| \leqq \epsilon_1 \|A_\alpha B_\beta f\|.$$

Once more, this choice of δ is uniform for all λ. Since

$$A_\lambda + B_\lambda = (1/2i)(\exp(-i\lambda)V - \exp(i\lambda)V^{-1}),$$

it is seen that, for $g=A_\alpha B_\beta f$, $\|(V^2-\exp(2i\lambda))^2 g\| \leqq 4K\epsilon_1\|g\|$. This inequality holds not only for the elements g but for all elements in the c.l.m. $\mathfrak{C}_{\lambda\delta}$ spanned by the g.

Applying (10) and what follows, it is possible to write for all $g \in \mathfrak{C}_{\lambda\delta}$

$$\|(V^2 - e^{2i\lambda})g\| \leqq \epsilon_2\|g\|$$

provided that $4K^2(n+1)\epsilon_1 \leqq \epsilon_2$ where the integer n is properly chosen as indicated above.

Let $g=g_1+g_2$ where the resolution is made as in (11) with V replaced by $\exp(-i\lambda)V$. Writing $(V+\exp(i\lambda))g_2=\zeta$, it is clear that $\|\zeta\| \leqq \epsilon_2\|g\|/2$. Henceforth assume that $\delta<\pi/2$. Simple considerations show that, for such a δ, $A_{\lambda-\delta}B_{\lambda+\delta}B_{\lambda-\pi/2}=0$ identically in $\mathfrak{B}$. Writing $B_\lambda=\sum_{-\infty}^{\infty}\alpha_s(\lambda)V^s$, it is seen that, for a given $\epsilon_3>0$, an integer $r\geqq 1$ may be found such that

$$K\sum_r^\infty |\alpha_s(\lambda)| + K\sum_{-\infty}^{-r} |\alpha_s(\lambda)| \leqq \epsilon_3$$

for all λ. Using the equations

$$V^s g_2 = (-1)^s e^{is\lambda} g_2 + V^{s-1}\zeta - e^{i\lambda}V^{s-2}\zeta + \cdots + (-1)^{s-1}e^{i(s-1)\lambda}\zeta, \qquad s = 0, 1, 2, \cdots,$$

$$V^{-s} g_2 = (-1)^s e^{-is\lambda} g_2 + e^{-i\lambda}V^{-s}\zeta - e^{-2i\lambda}V^{-s+1}\zeta + \cdots + (-1)^{s-1}e^{-is\lambda}V^{-1}\zeta, \qquad s = 0, 1, 2, \cdots,$$

and noting that if for a given $f \in \mathfrak{B}$ with $Vf=-\exp(i\lambda)f$, then $B_{\lambda-\pi/2}f=-f$, one may write

$$B_{\lambda-\pi/2}g_2 = -g_2 + \sum_r^\infty \alpha_s\left(\lambda - \frac{\pi}{2}\right)V^s g_2 + \sum_{-\infty}^{-r} \alpha_s\left(\lambda - \frac{\pi}{2}\right)V^s g_2$$
$$- \sum_r^\infty \alpha_s\left(-\frac{\pi}{2}\right)g_2 - \sum_{-\infty}^{-r} \alpha_s\left(-\frac{\pi}{2}\right)g_2 + \eta$$

where η stands for a collection of terms of the form $\beta V^s\zeta$ with $|\beta| \leqq 1$. Now the norm of the first two $\sum$ terms above does not exceed $\epsilon_3\|g_2\|$. The norm of the next two $\sum$ terms does not exceed $\epsilon_3\|g_2\|$. The term η represents a

sum of at most $r(r+1)$ terms each of norm less than $K\epsilon_2\|g\|/2$. Further, we have $B_{\lambda-\pi/2}g_2=\frac{1}{2}B_{\lambda-\pi/2}\cdot A_{\lambda-\delta}\cdot B_{\lambda+\delta}(f-e^{-i\lambda}Vf)=0$. Thus if it happens that $r(r+1)K\epsilon_2\leqq 2\epsilon_3$, then $\|g_2\|\leqq 2\epsilon_3\|g_2\|+\epsilon_3\|g\|$, or if $\epsilon_3<\frac{1}{4}$, $\|g_2\|\leqq 2\epsilon_3\|g\|$.

Finally, in the equation

$$(V-e^{i\lambda})(g_1+g_2)=e^{-i\lambda}\frac{V^2-e^{2i\lambda}}{2}g+(V-e^{i\lambda})g_2$$

it is seen that the norm of the first term after the equality symbol does not exceed $\epsilon_2\|g\|/2$. The norm of the last term does not exceed $4K\epsilon_3\|g\|$. Hence the norm of the term on the left of the equality sign does not exceed $5K\epsilon_3\|g\|$.

It is a simple matter now to compute δ for the given ϵ of the theorem. Choose ϵ_3 so that $5K\epsilon_3\leqq\epsilon$. Then choose as indicated above the numbers r, ϵ_2, n, ϵ_1 and finally δ (such that $\delta\leqq\epsilon$), in that order. For that choice, (12) holds for all elements in $\mathfrak{C}_{\lambda\delta}$.

The classes $\mathfrak{A}_\lambda$ will be defined. For $-\infty<\lambda<-\pi$, $\mathfrak{A}_\lambda=0$. For $\pi<\lambda<\infty$, $\mathfrak{A}_\lambda=0$. For $-\pi\leqq\lambda\leqq 0$, $\mathfrak{A}_\lambda$ is the intersection of the c.l.m. $\mathfrak{M}_0$ (see Theorem 7) with the c.l.m. $\mathfrak{C}_{\lambda\delta}$. For $0<\lambda\leqq\pi$, $\mathfrak{A}_\lambda$ is the intersection of the c.l.m. $\mathfrak{N}_0$ with the c.l.m. $\mathfrak{C}_{\lambda\delta}$.

Proof of (a) *in the theorem*. Consider specifically the case $-\pi<\lambda<0$; other cases, being similar to this one, will not be discussed. Let $g=A_{\lambda-\delta}B_{\lambda+\delta}f$. Since $g\ \varepsilon\ \mathfrak{M}_0$, $Ag=0$; also $A_{\lambda+\delta}g=0$. If $\lambda+\delta\geqq 0$, $g\ \varepsilon\ \mathfrak{E}_{\lambda+\epsilon}$, since $\mathfrak{M}_0=\mathfrak{E}_0\subset\mathfrak{E}_{\lambda+\delta}\subset\mathfrak{E}_{\lambda+\epsilon}$. If $\lambda+\delta<0$, let $f=g_0+g_0'+g_{\lambda+\delta}+g_{\lambda+\delta}'+h$ be the decomposition of f, (9), for $n=2$, $\lambda_1=0$, $\lambda_2=\lambda+\delta$. Then the decomposition of g is found by operating on the decomposition of f with $A_{\lambda-\delta}B_{\lambda+\delta}$. It is found that $B_{\lambda+\delta}g_0=0$, $A_{\lambda-\delta}g_{\lambda+\delta}'=0$, $A_{\lambda-\delta}B_{\lambda+\delta}g_0'=0$ (since $g\ \varepsilon\ \mathfrak{M}_0$). Thus $g\ \varepsilon\ \mathfrak{E}_{\lambda+\delta}\subset\mathfrak{E}_{\lambda+\epsilon}$. Hence $\mathfrak{A}_\lambda\subset\mathfrak{E}_{\lambda+\epsilon}$.

To prove that $g\ \varepsilon\ \mathfrak{F}_{\lambda-\epsilon}$, it suffices to show that $g\ \varepsilon\ \mathfrak{F}_{\lambda-\delta}$. Note first that $BB_{\lambda-\delta}g=0$. Using the decomposition for f, $f=g_0+g_0'+g_{\lambda-\delta}+g_{\lambda-\delta}'+h$, it is seen that $A_{\lambda-\delta}B_{\lambda+\delta}g_{\lambda-\delta}=0$, hence $g\ \varepsilon\ \mathfrak{F}_{\lambda-\delta}$. The proof of (b) falls out of the definition of $\mathfrak{A}_\lambda$.

The finite collection of manifolds $\mathfrak{A}_{\lambda_j}$, $j=1, 2, \cdots, n$, which is mentioned in the theorem will now be exhibited. Let $\lambda_1, \cdots, \lambda_n$, $-\pi<\lambda_j<\pi$, $\lambda_j\neq 0$, be any collection of numbers having the properties that (i) the open segments of length 2δ centered about the points $\exp(i\lambda_j)$, $j=1, \cdots, n$, on the complex unit circle cover that circle; and (ii) for every j, $j=1, \cdots, n$, there exists a k such that $\lambda_j=-\lambda_k$.

It will be proved that the $\mathfrak{A}_{\lambda_j}$, $j=1, \cdots, n$, together span $\mathfrak{B}$. Consider the set M of numbers: ± 1, $\pm\exp(i\lambda_j-i\delta)$, $\pm\exp(i\lambda_j+i\delta)$; these numbers may not be distinct. Let $\mu_1, \mu_2, \cdots, \mu_s$, $-\pi<\mu_j\leqq 0$, be distinct real numbers such that every number in the set N of numbers $\pm\exp(i\mu_j)$, $j=1, \cdots, s$, is found in M and conversely. Let $f=\sum(g_{\mu_j}+g_{\mu_j}')+h$ be the decomposition (9) of f which is generated by the μ_j. As before, h is in the c.l.m. spanned by the elements $k=\prod_{j=1}^{s}(A_{\mu_j}+B_{\mu_j})l$. Note that this c.l.m. is identical with that spanned by the elements $k=\prod_{j=1}^{s}(A_{\mu_j}+B_{\mu_j})^{t_j}l$ where the t_j are any positive

integers. In other words, k is in the c.l.m. spanned by the elements

$$k = (A + B)\prod_{j=1}^{n} (A_{\lambda_j-\delta} + B_{\lambda_j-\delta})(A_{\lambda_j+\delta} + B_{\lambda_j+\delta})l, \tag{13}$$

$l \varepsilon \mathfrak{B}$. It will be shown first that each g_{μ_j} and each g'_{μ_j} is in some $\mathfrak{A}_{\lambda_j}$; next that every term in the expansion of k in (13) is in some $\mathfrak{A}_{\lambda_j}$.

Consider any g_{μ_j} or g'_{μ_j}; for short call it g, and suppose for example that $Vg=\exp(i\lambda_j-i\delta)g$ (the case $Vg=\pm g$ is tacitly assumed to have been disposed of). By condition (i) on the λ_j, the point $\exp(i\lambda_j+i\delta)$ is an interior point of some interval centered about some λ_k. Hence $g \varepsilon \mathfrak{E}_{\lambda_k}\delta$. Invoking if necessary condition (ii) on the λ_j, it may be seen that g is in some $\mathfrak{A}_{\lambda_j}$.

The terms in the expansion of (13) may be written in the form

$$T = D\prod_{j=1}^{n} D_jE_j \tag{14}$$

where D is either A or B, D_j is either $A_{\lambda_j-\delta}$ or $B_{\lambda_j-\delta}$, and E_j is either $A_{\lambda_j+\delta}$ or $B_{\lambda_j+\delta}$. For some j, it is conceivable that D_j and E_j both represent an A; or again that both represent a B. Such a value of j will be described as a *resemblance value*. Suppose that D represents A. If all j with $0<\lambda_j<\pi$ (note that $-\pi$, 0, π are excluded as possible values for λ_j) are resemblance values, then by an argument now familiar, the transformation T (14) is identically 0 in $\mathfrak{B}$, and Tl certainly is in all $\mathfrak{A}_{\lambda_j}$. Suppose that a particular j is not a resemblance value. Then either the product $A_{\lambda_j-\delta}B_{\lambda_j+\delta}$ or the product $A_{\lambda_j+\delta}B_{\lambda_j-\delta}$ is present in (14). In the first case $Tl \varepsilon A_{\lambda_j}$. If on the other hand the second case arises for all non-resemblance values j, then once more $Tl=0$. If D represents B, the discussion is similar. This concludes the proof of the theorem.

V. THE SPECTRUM OF V

The theory so far elaborated enables one to discuss the spectrum of any w.a.p. transformation V. One defines as usual the point spectrum, continuous spectrum, and resolvent set for V: A complex number α belongs to the point spectrum if $Vf=\alpha f$ for some $f\neq 0$. The number α belongs to the continuous spectrum if the transformation $V-\alpha I$ possesses an unbounded inverse whose domain is dense in $\mathfrak{B}$. Finally the value α belongs to the resolvent set if the transformation $V-\alpha I$ possesses a bounded inverse whose domain is $\mathfrak{B}$ in its entirety. The above classes are mutually exclusive but are not for a general (non-w.a.p.) transformation all inclusive. It will be shown below that they cover all the possibilities for V.

Relative to the manifold pairs $\{\mathfrak{E}_\lambda, \mathfrak{F}_\lambda\}$, one may define certain classes of numbers in the following fashion: The numbers λ for which there exists a $\delta>0$ such that $\mathfrak{E}_{\lambda-\delta}=\mathfrak{E}_{\lambda+\delta}$, $\mathfrak{F}_{\lambda-\delta}=\mathfrak{F}_{\lambda+\delta}$ constitute *the points of constancy* of $\{\mathfrak{E}_\lambda, \mathfrak{F}_\lambda\}$. For instance the numbers $\lambda<-\pi$ and $\lambda>\pi$ are points of constancy. The num-

bers λ which are not points of constancy but for which $\mathfrak{F}_{\lambda-0}=\mathfrak{F}_\lambda$ will constitute the *points of continuity* of $\{\mathfrak{E}_\lambda, \mathfrak{F}_\lambda\}$ [12]. The numbers λ for which $\mathfrak{F}_{\lambda-0}\neq\mathfrak{F}_\lambda$ will constitute the *points of discontinuity* of $\{\mathfrak{E}_\lambda, \mathfrak{F}_\lambda\}$. Thus each real λ belongs to one and only one of these three types. The striking relationship of the character of $\{\mathfrak{E}_\lambda, \mathfrak{F}_\lambda\}$ to the spectrum of V is exposed in the following theorem.

THEOREM 9. *Let V be a weakly almost-periodic transformation, and let $\{\mathfrak{E}_\lambda, \mathfrak{F}_\lambda\}$ be the manifold pairs associated with it in Theorem 7. Let α be any complex number. Then if $|\alpha|\neq 1$, α belongs to the resolvent set of V. If the real number λ is restricted to the interval $-\pi<\lambda<\pi$, then*

(a) *The value $\alpha=\exp(i\lambda)$ belongs to the point spectrum if and only if λ is a point of discontinuity of $\{\mathfrak{E}_\lambda, \mathfrak{F}_\lambda\}$.*

(b) *The value $\alpha=\exp(i\lambda)$ belongs to the continuous spectrum if and only if λ is a point of continuity of $\{\mathfrak{E}_\lambda, \mathfrak{F}_\lambda\}$.*

(c) *The value $\alpha=\exp(i\lambda)$ belongs to the resolvent set if and only if λ is a point of constancy of $\{\mathfrak{E}_\lambda, \mathfrak{F}_\lambda\}$.*

The value $\alpha=\exp(-i\pi)=\exp(i\pi)$ belongs to the resolvent set if and only if $-\pi$ and π are points of constancy of $\{\mathfrak{E}_\lambda, \mathfrak{F}_\lambda\}$. The value $\exp(i\pi)$ *belongs to the point spectrum if and only if π is a point of discontinuity of $\{\mathfrak{E}_\lambda, \mathfrak{F}_\lambda\}$. In all other cases,* $\exp(i\pi)$ *belongs to the continuous spectrum.*

Suppose first that $|\alpha|>1$. Then $|(\alpha^{-1}V)^n|\leqq|\alpha|^{-n}K$ where as before $|V^n|\leqq K$. Thus by Theorem 4′, the transformation $\alpha^{-1}V-I$ possesses a bounded inverse; hence $V-\alpha I$ does likewise.

Now let $|\alpha|<1$. By the reasoning just given $V^{-1}(V-\alpha I)=I-\alpha V^{-1}$ possesses a bounded inverse. Hence $V-\alpha I$ does likewise.

If $|\alpha|=1$, then by Theorem 4. if $(\alpha^{-1}V-I)f=0$ has no non-trivial solutions, the elements $(\alpha^{-1}V-I)f$, $f\in\mathfrak{B}$, are dense in $\mathfrak{B}$. Thus if α does not belong to the point spectrum of V, it belongs either to the continuous spectrum or to the resolvent set. For if the range of $S=\alpha^{-1}V-I$ is $\mathfrak{B}$ in its entirety, S gives a one-to-one mapping of $\mathfrak{B}$ into itself which is continuous in one direction; therefore the mapping is continuous in the reverse direction. In this case α belongs to the resolvent set. If the range of S is dense in $\mathfrak{B}$ but is not $\mathfrak{B}$, then S^{-1} cannot be bounded; for if S^{-1} were bounded, the range of S could be extended to the entire space $\mathfrak{B}$. In this case α is in the continuous spectrum.

Toward (a). Suppose $\alpha=\exp(i\lambda)$ belongs to the point spectrum and that $Vf=\exp(i\lambda)f$, $f\neq 0$. Then it is easy to conclude that $f\in\mathfrak{F}_\mu$, $\mu<\lambda$. On the other hand $f\in\mathfrak{E}_\lambda$, hence f is not in $\mathfrak{F}_\lambda$. Thus $\mathfrak{F}_{\lambda-0}\neq\mathfrak{F}_\lambda$.

Let $f\in\mathfrak{F}_{\lambda-0}$, $f\notin\mathfrak{F}_\lambda$; and let $f=g_0+g_0'+g_\lambda+g_\lambda'+h$ be a partition (9) of f. Assume for the sake of definiteness that $-\pi<\lambda<0$. Since $BB_\mu f=0$, $-\pi<\mu<\lambda$, $BB_\lambda f=0$ by virtue of the continuity of BB_μ. Since $f\notin\mathfrak{F}_\lambda$, $g_\lambda\neq 0$. Thus there exists an element $g_\lambda\neq 0$ such that $Vg_\lambda=\alpha g_\lambda$ and α is in the point spectrum of V.

---

[12] $\mathfrak{F}_{\lambda-0}$ shall be defined to be the intersection of all $\mathfrak{F}_\mu$ with $\mu<\lambda$: $\mathfrak{F}_{\lambda-0}=\prod_{\mu<\lambda}\mathfrak{F}_\mu$.

Toward (b). Let $\exp(i\lambda)$ be in the continuous spectrum. It will be shown that for any $\delta>0$ with $-\pi<\lambda-\delta<\lambda+\delta<\pi$ there exists an element $g=A_{\lambda-\delta}B_{\lambda+\delta}f\neq 0$. Since $g\ \varepsilon\ \mathfrak{E}_{\lambda+\delta}$ and $g\ \varepsilon\ \mathfrak{F}_{\lambda-\delta}$, the equation $\mathfrak{E}_{\lambda+\delta}=\mathfrak{E}_{\lambda-\delta}$ is impossible; thus λ is not a point of constancy of $\{\mathfrak{E}_\lambda, \mathfrak{F}_\lambda\}$. Since λ is not a point of discontinuity by (a), λ must be a point of continuity of $\{\mathfrak{E}_\lambda, \mathfrak{F}_\lambda\}$.

Since $\exp(i\lambda)$ is in the continuous spectrum, it is possible, given an integer N and an $\epsilon>0$, to find an $f\neq 0$ such that $\|(V^n-\exp(in\lambda))f\|\leqq\epsilon\|f\|$, $n=\pm 1, \pm 2, \cdots, \pm N$. Thus for such an f the elements

$$(A_{\lambda-\delta}-\sin\delta)f,\quad A_{\lambda+\delta}f,\qquad (B_{\lambda+\delta}+\sin\delta)f,\quad B_{\lambda-\delta}f,$$

have a norm small at will compared to that of f. This is proved as in Theorem 8 in the discussion preceding (a). Thus an $f\ \varepsilon\ \mathfrak{B}$ can be found such that $A_{\lambda-\delta}B_{\lambda+\delta}f+\sin^2\delta f$ has a norm small at will compared to that of f. This means that $A_{\lambda-\delta}B_{\lambda+\delta}f\neq 0$.

In proving the converse, one starts with the fact that the simultaneous equations $\mathfrak{E}_{\lambda+\delta}=\mathfrak{E}_{\lambda-\delta}$, $\mathfrak{F}_{\lambda+\delta}=\mathfrak{F}_{\lambda-\delta}$ are impossible for any $\delta>0$. Assume that the former equation does not hold, hence that there exists an $f\neq 0$ such that $f\ \varepsilon\ \mathfrak{E}_{\lambda+\delta}$, $f\ \not\varepsilon\ \mathfrak{E}_{\lambda-\delta}$. Assume for immediate convenience that $-\pi<\lambda-2\delta<\lambda+2\delta<0$. If $g=A_{\lambda-2\delta}f=0$, $f\ \varepsilon\ \mathfrak{E}_{\lambda-\delta}$ which is impossible. If $B_{\lambda+2\delta}g=0$, $g\ \varepsilon\ \mathfrak{F}_{\lambda+\delta}$ which is impossible since, like f, $g\ \varepsilon\ \mathfrak{E}_{\lambda+\delta}$. Since δ is arbitrary, by Theorem 8, (12), $V-\exp(i\lambda)I$ does not possess a bounded inverse, hence $\exp(i\lambda)$ is either in the continuous spectrum or in the point spectrum. The latter possibility is barred by the proof of (a) of the present theorem. This demonstration has been carried through for a sample case; other cases may be handled in a similar manner.

Toward (c). Proof here is obtained by applying the exclusion principle. The statements in the theorem concerning the value $\alpha=\exp(-i\pi)=\exp(i\pi)$ will not be discussed as the methods of proof here needed are not very different from those already indicated.

VI. The transformation H

Throughout the present chapter, V will be a w.a.p. transformation for which the transformation $(V+I)^{-1}$ exists, that is, for which $\alpha=-1$ is not in the point spectrum. A very important function, H, of V is defined by

$$H=-i(V-I)(V+I)^{-1}. \tag{15}$$

The meaning of (15) is to be precisely this: H is defined for all elements f of the form $f=(V+I)g$, g arbitrary in $\mathfrak{B}$; and $Hf=-i(V-I)g$. The transformation H enjoys closure. Let the sequence $\{f_n\}$ be in the domain of H. Let $f_n\to f$, and $Hf_n\to f'$. Then f is in the domain of H and $Hf=f'$. For writing $f_n=(V+I)g_n$ and $Hf_n=-i(V-I)g_n$, it is clear that the convergence of $\{f_n\}$ and of $\{Hf_n\}$ implies that of $\{g_n\}$. Let $g_n\to g$, then $f=(V+I)g$ and $f'=\lim[-i(V-I)g_n]=-i(V-I)g=Hf$. If $(V+I)^{-1}$ is bounded, then H is

bounded. Conversely, if H is bounded, since it is closed and defined over a dense set in $\mathfrak{B}$, its domain of definition is $\mathfrak{B}$ in its entirety; thus $(V+I)^{-1}$ is bounded. Thus H is bounded if and only if there exists a $\lambda<\pi$ such that $\mathfrak{E}_\lambda=\mathfrak{B}$, $\mathfrak{F}_\lambda=0$, $\mathfrak{E}_{-\lambda}=0$, $\mathfrak{F}_{-\lambda}=\mathfrak{B}$.

Let λ and μ be real variables related by $\lambda=\tan\mu/2$, $-\pi<\mu<\pi$. Consider the manifold pairs $\{\mathfrak{E}_\lambda, \mathfrak{F}_\lambda\}$ of Theorem 7 with λ restricted to $-\pi<\lambda<\pi$. Construct the pairs of manifolds $\{\mathfrak{E}_\lambda', \mathfrak{F}_\lambda'\}$, $-\infty<\lambda<\infty$, where $\mathfrak{E}_\lambda'=\mathfrak{E}_\mu$, $\mathfrak{F}_\lambda'=\mathfrak{F}_\mu$. Then all the statements of Theorem 7 except (d) are valid if $\{\mathfrak{E}_\lambda, \mathfrak{F}_\lambda\}$ is replaced by $\{\mathfrak{E}_\lambda', \mathfrak{F}_\lambda'\}$. In place of (d), one may write $\prod_\lambda\mathfrak{F}_\lambda'=0$, $\prod_\lambda\mathfrak{E}_\lambda'=0$, where the product is formed over all real λ. The first formula is established with the help of the fact that $\mathfrak{F}_{\pi-0}=\mathfrak{F}_\pi=0$. The second may be verified readily.

The point spectrum, continuous spectrum, and resolvent set of H are defined as for V (Chapter V). Reducibility of H by a manifold pair $\{\mathfrak{M}, \mathfrak{N}\}$ could be defined as for V, but the appearance of unboundedness creates difficulties which lead to the abandonment of this idea. Instead, the pair $\{\mathfrak{M}, \mathfrak{N}\}$ will be said to reduce H if it reduces V. If $\mathfrak{M}$ and $\mathfrak{N}$ are disjoint, this implies that Hf is in $\mathfrak{M}$ (or $\mathfrak{N}$) whenever f is in $\mathfrak{M}$ (or $\mathfrak{N}$). For let $f\,\varepsilon\,\mathfrak{M}$, $f=(V+I)g$, $g=g_1+g_2$ with $g_1\,\varepsilon\,\mathfrak{M}$, $g_2\,\varepsilon\,\mathfrak{N}$. Then since $\{\mathfrak{M}, \mathfrak{N}\}$ reduces V, $(V+I)g_2\,\varepsilon\,\mathfrak{N}$; but $(V+I)g_2=f-(V+I)g_1\,\varepsilon\,\mathfrak{M}$. Thus $(V+I)g_2=0$, $g_2=0$. Hence $Hf\,\varepsilon\,\mathfrak{M}$.

The relationship of $\{\mathfrak{E}_\lambda', \mathfrak{F}_\lambda'\}$ to H is brought to light in the next two theorems. The first establishes for H an integral representation (1).

THEOREM 10. *Let H be the transformation $H=-i(V-I)(V+I)^{-1}$ and let $\{\mathfrak{E}_\lambda', \mathfrak{F}_\lambda'\}$, $-\infty<\lambda<\infty$, be the manifold pairs constructed above. Let a real number $\epsilon>0$ be given. Then for every real λ there exists a closed linear manifold $\mathfrak{D}_\lambda$ such that:*

(a) *The manifold $\mathfrak{D}_\lambda$ lies in $\mathfrak{E}_{\lambda+\epsilon}'$ and $\mathfrak{F}_{\lambda-\epsilon}'$.*

(b) *If $f\,\varepsilon\,\mathfrak{D}_\lambda$, then Hf exists and*

$$\|(H-\lambda I)f\|\leqq\epsilon\|f\|. \tag{16}$$

Furthermore, there exists a denumerable collection of values λ, viz., $\lambda_1, \lambda_2, \cdots$, such that the space $\mathfrak{B}$ is spanned by the manifolds $\mathfrak{D}_{\lambda_j}$, $j=1, 2, \cdots$. If H is bounded, this infinite collection of λ_j may be replaced by a finite set.

The definition of the c.l.m. $\mathfrak{D}_\lambda$ must be prefaced by a brief discussion. Consider first the c.l.m. $\mathfrak{A}_\mu$ of Theorem 8 for a μ such that $-\pi<\mu<\pi$. Specifically, let $\mathfrak{A}_\mu$ satisfy the conditions of Theorem 8 relative to an $\epsilon_1<2\cos\mu/2$; further let the δ of $\mathfrak{A}_\mu$ satisfy $-\pi<\mu-\delta<\mu+\delta<\pi$. Then the transformation $V+I$ transforms $\mathfrak{A}_\mu$ into itself in a one-to-one bicontinuous manner. This will be proved by showing first that the range of $V+I$ (defined over $\mathfrak{A}_\mu$) is a c.l.m.; subsequently that this range is dense in $\mathfrak{A}_\mu$.

By the definition of $\mathfrak{A}_\mu$, $(V+I)g\,\varepsilon\,\mathfrak{A}_\mu$ if $g\,\varepsilon\,\mathfrak{A}_\mu$. Let $\{g_n\}$ be a sequence in $\mathfrak{A}_\mu$, let $f_n=(V+I)g_n$, and suppose that $f_n\to f$. Writing $Vg_n=\exp(i\mu)g_n+\eta_n$, it follows from (12) that $\|\eta_n\|\leqq\epsilon_1\|g_n\|$, $\|\eta_n-\eta_m\|\leqq\epsilon_1\|g_n-g_m\|$. Thus $\|g_n-g_m\|$

$\leqq(2\cos\mu/2-\epsilon_1)^{-1}\|f_n-f_m\|$, $\{g_n\}$ is convergent, say $g_n\to g$ and $(V+I)g=f$. Hence the range of $V+I$ is a c.l.m.

Now let g be arbitrary in $\mathfrak{A}_\mu$. Using the identity $(V+I)g=(\exp(i\mu)+1)g+(V-\exp(i\mu))g$, and writing $\alpha=(\exp(i\mu)+1)^{-1}$, it is seen that

$$(V+I)[\alpha I-\alpha^2(V-e^{i\mu})+\alpha^3(V-e^{i\mu})^2-\cdots+(-1)^{n+1}\alpha^n(V-e^{i\mu})^{n-1}]g$$
$$=g+(-1)^{n+1}\alpha^n(V-e^{i\mu})^n g.$$

The norm of this last element does not exceed $\epsilon_1^n(2\cos\mu/2)^{-n}\|g\|$, hence is small at will if n is sufficiently large. This proves that the range of $V+I$ is $\mathfrak{A}_\mu$.

The transformation H is defined and bounded on $\mathfrak{A}_\mu$. To compute the norm on $\mathfrak{A}_\mu$ of $H-\lambda I$ where $\lambda=\tan\mu/2$, write $Hf=-i(V-I)g$ with $(V+I)g=f$. Calculation shows that $(H-\lambda I)f=-(i+\lambda)[V-\exp(i\mu)]g$. Since $\|g\|\leqq(2\cos\mu/2-\epsilon_1)^{-1}\|f\|$, $|H-\lambda I|\leqq\epsilon$ if $\epsilon_1(1+|\lambda|)(2\cos\mu/2-\epsilon_1)^{-1}\leqq\epsilon$, where ϵ is the constant mentioned in the statement of the present theorem.

The definition of $\mathfrak{D}_\lambda$ is: $\mathfrak{D}_\lambda=\mathfrak{A}_\mu$, $\lambda=\tan\mu/2$. Note that (b) has been established. It was demonstrated in Theorem 8 that $\mathfrak{A}_\mu$ lies in $\mathfrak{E}_{\mu+\delta}$ as well as in $\mathfrak{F}_{\mu-\delta}$. Thus $\mathfrak{D}_\lambda$ lies in $\mathfrak{E}'_{\lambda_1}$, $\mathfrak{F}'_{\lambda_2}$, where $\lambda_1=\tan(\mu+\delta)/2$, $\lambda_2=\tan(\mu-\delta)/2$. By choosing δ properly, one may make $\lambda_1\leqq\lambda+\epsilon$, $\lambda_2\geqq\lambda-\epsilon$. This proves (a).

The collection of values $\lambda_1,\lambda_2,\cdots$ will be exhibited. The procedure just outlined for the choice of $\mathfrak{D}_\lambda=\mathfrak{A}_\mu$ embeds every point $\exp(i\mu)$ on the complex unit circle (except the point $\mu=\pi$) in an interval of varying length 2δ, δ depending on μ. Let $\mu_1,\mu_2,\cdots,-\pi<\mu_j<\pi,\mu_j\neq0$, be any sequence of numbers having the properties: (i) The open intervals of length 2δ centered about $\exp(i\mu)$ cover completely the unit circle from which the point $\mu=\pi$ has been excluded; (ii) for every μ_j there exists a μ_k such that $\mu_j=-\mu_k$. The sequence $\lambda_1,\lambda_2,\cdots$ is defined by $\lambda_j=\tan\mu_j/2$, $j=1,2,\cdots$.

Let F be any linear functional in $(\mathfrak{B})$ such that $F\perp\mathfrak{D}_{\lambda_j}$, $j=1,2,\cdots$. It will be proved that $F=0$, hence that the c.l.m. spanned by the $\mathfrak{D}_{\lambda_j}$ is $\mathfrak{B}$. Let ν be any number, $-\pi<\nu<0$. Let n be an integer such that the semicircle $\exp(ix)$, $\nu\leqq x\leqq\nu+\pi$ is covered by the above described intervals centered about the points $\exp(i\mu_j)$, $j=1,2,\cdots,n$. Then the manifolds $\mathfrak{M}_\nu$, $\mathfrak{A}_{\mu_j}$, $j=1,\cdots,n$, span $\mathfrak{B}$(13). Now for $\overline{A}_\nu F$ ($\overline{A}_\nu$ is the transformation adjoint to A_ν), $\overline{A}_\nu F\perp\mathfrak{M}_\nu$ as well as $\overline{A}_\nu F\perp\mathfrak{A}_{\mu_j}$, $j=1,\cdots,n$. Hence $\overline{A}_\nu F=0$. Noting that ν is arbitrary, and letting $\nu\to0$, one has $\overline{A}F=0$. A very similar argument using a ν with $0<\nu<\pi$ shows that $\overline{B}F=0$. Thus for F, $\overline{V}F=-F$. By Theorem 3, the collection of elements F having the latter property is the orthogonal complement of the c.l.m. $\mathfrak{B}$ spanned by the elements $(V+I)f$. Therefore $F=0$.

If H is bounded, the points $\mu=-\pi$ and $\mu=\pi$ are points of constancy of $\{\mathfrak{E}_\mu,\mathfrak{F}_\mu\}$ and it is easy to see that only a finite number of the manifolds $\mathfrak{A}_{\mu_j}$ contribute effectively to the spanning of $\mathfrak{B}$.

(13) The argument needed to establish this is similar in spirit to that given in the last paragraphs of Chapter IV. To the collection of transformations $A_{\mu_j\pm\delta}$, $B_{\mu_j\pm\delta}$, A, and B, one must adjoin A_ν, B_ν.

In the next theorem, the spectral characteristics of H are completely described by the structure of $\{\mathfrak{E}'_\lambda, \mathfrak{F}'_\lambda\}$.

THEOREM 11. *Let H and $\{\mathfrak{E}'_\lambda, \mathfrak{F}'_\lambda\}$ be as in Theorem 10. Let $\alpha=\lambda+\lambda' i$ be any complex number. The value α is in the resolvent set of H whenever $\lambda'\neq 0$. The transformation V which generates H is completely determined by H:*

$$V = -(H - iI)(H + iI)^{-1}. \tag{17}$$

If $\lambda'=0$, then

(a) *The value $\alpha=\lambda$ belongs to the point spectrum of H if and only if λ is a point of discontinuity of $\{\mathfrak{E}'_\lambda, \mathfrak{F}'_\lambda\}$.*

(b) *The value $\alpha=\lambda$ belongs to the continuous spectrum of H if and only if λ is a point of continuity of $\{\mathfrak{E}'_\lambda, \mathfrak{F}'_\lambda\}$.*

(c) *The value $\alpha=\lambda$ belongs to the resolvent set of H if and only if λ is a point of constancy of $\{\mathfrak{E}'_\lambda, \mathfrak{F}'_\lambda\}$.*

Write $f=(V+I)g$, $H_\alpha f=(H-\alpha I)f=-(i+\alpha)Vg+(i-\alpha)g$. If $\alpha=-i$, $H_{-i}f=0$ implies $f=0$. The range of H_{-i} is $\mathfrak{B}$; if $\{H_{-i}f_n\}$ is convergent, then $\{f_n\}$ is convergent. Thus $\alpha=-i$ is in the resolvent set of H. Henceforth let $\alpha\neq -i$. Then $H_\alpha f=-(i+\alpha)(V-\beta I)g$, where $\beta=(i-\alpha)(i+\alpha)^{-1}$. If $\lambda'\neq 0$, then $|\beta|\neq 1$. In this case $H_\alpha f=0$ implies $g=f=0$ by Theorem 9. The range of H_α is again $\mathfrak{B}$ since β is in the resolvent set of V. For the same reason, the convergence of $\{H_\alpha f_n\}$ with $f_n=(V+I)g_n$ implies that of $\{g_n\}$ and hence that of $\{f_n\}$; thus α is in the resolvent set of H.

Using the above notation, it is seen that $(H+iI)^{-1}g=(2i)^{-1}f$; $(H-iI)f=-2iVg$. This establishes (17).

Now let $\lambda'=0$, $\alpha=\lambda$; then $|\beta|=1$, or more precisely $\beta=\exp(i\mu)$ with $\tan \mu/2=\lambda$. Suppose $H_\lambda f=0$; then $Vg=\beta g$—and conversely. Thus λ is in the point spectrum of H if and only if β is in the point spectrum of V. Note that the range of H_λ is dense in $\mathfrak{B}$ if and only if λ is not in the point spectrum of H. Suppose the range of H_λ is dense in $\mathfrak{B}$ but is not $\mathfrak{B}$. Since H_λ is a closed transformation, $(H_\lambda)^{-1}$ cannot be bounded; thus λ is in the continuous spectrum of H. In this case, β is in the continuous spectrum of V. Conversely, if β is in the continuous spectrum of V, λ is in that of H. Finally, if the range of H_λ is $\mathfrak{B}$, then β is in the resolvent set of V and the convergence of $\{H_\lambda f_n\}$ implies that of $\{g_n\}$ and $\{f_n\}$. Thus if the range of H_λ is $\mathfrak{B}$, λ is in the resolvent set of H. Hence λ is in the resolvent set of H if and only if β is in the resolvent set of V.

It appears that the spectral character of H_λ is identical with that of $V_\mu-I$. Since the relation between λ and μ is $\lambda=\tan \mu/2$, and since $\{\mathfrak{E}'_\lambda, \mathfrak{F}'_\lambda\}$ and $\{\mathfrak{E}_\mu, \mathfrak{F}_\mu\}$ are related in the same way, the verification of statements (a), (b), and (c) is shouldered by Theorem 9.

BARNARD COLLEGE, COLUMBIA UNIVERSITY,
NEW YORK, N. Y.

GENERAL COMBINATORIAL TOPOLOGY

BY

PAUL ALEXANDROFF

To Serge Bernstein on his sixtieth birthday

After the fundamental conceptions of the so-called combinatorial topology were transferred by the author of the present paper(¹) as well as by Vietoris, Lefshetz, Čech and others to arbitrary compact metric spaces and, having obtained in the general duality law of Alexander-Pontrjagin, the homological theory of dimensionality and a number of other essentially new investigations a concrete geometrical development, became a mighty and generally recognized weapon in the investigation of different topological questions, it became not only possible to speak of a new branch of topology—*the homological theory of spaces*—but it also seemed that the directions of further development of this new branch were more or less determined. This latter opinion, however, was not confirmed: in 1934 Kolmogoroff(¹) and nearly simultaneously with him Alexander(²) gave to the development of the homological topology an essentially new direction by the discovery of the so-called upper boundary operator (which we call here the ∇-operator) dual to the old boundary operator (we call it here the Δ-operator) and permitting one to construct two systems of homological invariants dual to each other in the sense of the Pontrjagin theory of characters not only for polyhedrons and complexes but also for arbitrary locally bicompact spaces.

The central fact of the theory is a proposition which was originally formulated by Kolmogoroff and which we therefore call in the present paper *the duality law of Kolmogoroff*. This proposition asserts that for any closed set A lying in a locally bicompact space R, the r- and $(r+1)$-dimensional Betti groups of which are null groups, there exists an isomorphism between the r-dimensional Betti group of A and the $(r+1)$-dimensional Betti group of $R-A$. This proposition enables us to give a new meaning to the duality law

Presented to the Society, September 12, 1940; received by the editors April 6, 1940.

(¹) Kolmogoroff, International Tensor Conference and International Topological Conference, Moscow, May, 1934, and September, 1935; papers: (a) *Über die Dualität im Aufbau der kombinatorischen Topologie*, Recueil Mathématique de Moscou, vol. 1 (43) (1936), pp. 97–102; (b) *Les groupes de Betti des espaces localement bicompacts*; *Propriétés des groupes de Betti des espaces localement bicompacts*; *Les groupes de Betti des espaces metriques*; *Cycles relatifs, théorème de dualité de M. Alexander*—all four papers in the Comptes Rendus de l'Académie des Sciences, Paris, vol. 202 (1936), pp. 1144, 1325, 1558, 1641.

(²) Alexander (abstracted in the proceedings of the National Academy of Sciences for 1935): (a) *On the connectivity ring of an abstract space*, Annals of Mathematics, (2), vol. 37 (1936), pp. 698–708. (b) *A theory of connectivity in terms of gratings*, Annals of Mathematics, (2), vol. 39 (1938), pp. 883–912. The same ideas appear in a different form in the "pseudocycles" of Lefschetz (see his *Topology*) as soon as 1930.

of Alexander-Pontrjagin (which may be easily deduced from the duality law of Kolmogoroff under the assumption that R is the n-dimensional euclidean or spherical space).

In the construction of the theory itself Alexander and Kolmogoroff proceed differently. The construction of Kolmogoroff, a concise exposition of which without fundamental proofs was given by him in four notes in the Comptes Rendus de l'Académie des Sciences, Paris(1), is based on a completely new approach to homological problems of the set-theoretical topology and starts from the consideration of the functions $\phi^r(E_0, E_1, \cdots, E_r)$ and $f^r(e_0, e_1, \cdots, e_r)$, where the E_i are sets and the e_i-points of the given space. The functions $\phi^r(E_0, E_1, \cdots, E_r)$ are skew-symmetrical and finitely additive with respect to all their arguments; their values belong to the bicompact commutative group Ξ (the "field of coefficients"). The functions $f^r(e_0, e_1, \cdots, e_r)$ are also skew-symmetrical, but their values are taken from a discrete commutative group. The functions $\phi^r(E_0, E_1, \cdots, E_r)$ play the role of algebraical complexes of the usual combinatorial topology of complexes and are the starting point of the Δ-theory. As analogues of algebraical complexes in the ∇-theory appear not the functions $f^r(e_0, e_1, \cdots, e_r)$ themselves but classes of such functions equivalent to each other in a certain sense.

This way of construction may prove to be the most fruitful from the point of view of further investigations. But it considerably differs from the methods based on the elementary devices of combinatorial topology, which have dominated so far. This newness of the method as well as, undoubtedly, the fact that Kolmogoroff has not as yet given an exposition of the theory which is to any extent complete, nor, in particular, the proof of his duality law account, probably, for the fact that his theory is not yet as widespread as it deserves to be and that it has not so far influenced the further development of the topology to an extent to which it will doubtlessly influence it in the future.

Practically speaking, Alexander has realized his construction of the same theory in several different ways. A proof of the already mentioned duality law he gives, however, only in his last publication, *A theory of connectivity in terms of gratings*(2), where to this end the whole theory is constructed on an entirely new basis with the help of the so-called gratings. The equivalence of this theory with the other theories of Alexander (as well as with the theory of Kolmogoroff) is not yet proved, although it is highly probable. The apparatus of gratings applied by Alexander in his last paper has a very simple geometrical figure (the decomposition of the space by a plane into two half-spaces) for its source. But in the general setting in which the construction proceeds this original figure becomes so complicated that the whole resulting structure is extremely involved.

In the present paper the theory is built on a completely elementary basis, namely by means of well known considerations of the finite coverings of the given space. I left my old devices of application of combinatorial methods

to the study of general spaces only in one respect: along with the nerves of the coverings I consider now the ***barycentrical subdivisions of the coverings*** introduced by me recently elsewhere([3]), which, as I think, give us often a more elastic weapon for the study of topological properties of the given space.

The distribution of the material is as follows. In the first two sections we construct the Δ- and the ∇-theories for complexes. Here we systematize and prove things which are in the majority of cases known, but the proofs of which (and sometimes even the formulations) are, in a large proportion of cases, nowhere published([4]).

In §3 we recollect the notion of inverse and direct series of homomorphisms (instead of which we shall speak of direct and inverse spectra) of, in the general case partially ordered (not necessarily enumerable), systems of groups. This theory was originally constructed for enumerable sequences by Pontrjagin([5]) and for arbitrary systems by Steenrod([4]). I may point out the formulation of the conception of the limit group of a direct spectrum, which is logically simpler than the usual one.

In §4 we formulate and prove the "formal duality law" in application to arbitrary partially ordered systems of complexes. Substantially it is this "formal" duality that forms the combinatorial basis of the duality law of Kolmogoroff. The general formulation given here may prove convenient for application to the study of different concrete problems (local properties of sets, etc.).

In §§5 and 6 we prove the fundamental lemmas which shall be used in the proof of the duality law of Kolmogoroff. The distribution into two sections is made according to whether the lemmas concern finite coverings of an arbitrary set or special coverings of topological spaces. Thus the notion of the topological space we meet in the present paper for the first time in §6.

In §7 we give the first definition of Betti groups for any spaces homeomorphic to open sets lying in normal spaces. This definition hangs together with my old papers as well as with the paper of Steenrod referred to above, i.e., it defines the Betti groups of a space as limit groups of respectively the direct and the inverse spectra composed of Betti groups of the nerves of finite coverings of the given space by its open sets. For the sequel it is, however, of importance that along with the Betti groups which for normal spaces were defined by Steenrod, we define also other groups taking particularly into account those elements of the covering, the closures of which are bicompact.

([3]) Alexandroff, ***Diskrete Räume***, Recueil Mathématique de Moscou, vol. 2 (44) (1937), pp. 501–518.

([4]) See, however, Whitney, ***On matrices of integers and combinatorial topology***, Duke Mathematical Journal, vol. 3 (1937), and ***On products in a complex***, Annals of Mathematics, (2), vol. 39 (1938), pp. 397–432, as well as Steenrod, ***Universal homology groups***, American Journal of Mathematics, vol. 58 (1936), pp. 661–701.

([5]) Pontrjagin, ***Über den algebraischen Inhalt topologischer Dualitätssätze***, Mathematische Annalen, vol. 105 (1931), pp. 165–205.

It is just these groups, for the first time introduced in the present paper under the name of ***inner Betti groups***, that form the object of the duality law of Kolmogoroff.

The same §7 contains the proof of a theorem (Theorem 7.41), in which is formulated all that I know in the direction of the duality law of Kolmogoroff in the case when the space R is not locally bicompact.

In §8 we prove the duality law of Kolmogoroff for any locally bicompact normal space and closed $A \subset R$.

In §9 is given a new definition of inner Betti groups based on the consideration of barycentrical subdivisions of finite coverings of R by open sets and is proved the equivalence of this definition with the old one. At the end of this paragraph we give a formulation of the conception of Betti groups which does not use any auxiliary conceptions except the conception of the finite covering and the spectrum (series of homomorphisms) of groups.

In §10 is introduced the operation of multiplication of elements of Betti groups and in this way the complete Betti group (i.e., the direct sum of Betti groups of different dimensionalities) is turned into a ring (the connectivity ring). Although this definition of multiplication follows that of Alexander[6], it has in comparison with the latter an advantage consisting in the freedom from any special ordering of vertices which was applied in Alexander's paper *On the connectivity ring of an abstract space*[2], as well as in the freedom from any special conditions by means of which the multiplication is introduced in the paper *A theory of connectivity in terms of gratings*[2].

Notations

Throughout this paper the following notations are of constant use:

(i) $A \cup B$ means the set theoretical sum, $A \cap B$ means the set theoretical intersection of the sets A and B.

(ii) $\bigcup_\alpha A_\alpha$ means the set theoretical sum and $\bigcap_\alpha A_\alpha$ the intersection of all sets A_α of a given family of sets.

(iii) $A - B$ means the difference between the set A and the set B, i.e., the set of all elements of A not belonging to B. In this way

$$A - B = A - A \cap B,$$

where is it not supposed that B is a subset of A. But if A is a commutative group and B a subgroup of A, then $A - B$ means the factor (or difference) group of A with respect to B.

(iv) $A \subset B$ means that every element of the set A is an element of the set B (the identity $A = B$ being not excluded).

(v) $a \in A$ means that a is an element of the set A.

---

(6) Introduced in the paper *On the connectivity ring of an abstract space* (see footnote 2). See also in that paper the references to Čech and Whitney.

(vi) If t and t' are simplexes, then $t' < t$ means that t' is a face of t. The letter t denotes a non-oriented as well as an oriented simplex.

1. Complexes

1.1. By a *complex* we shall mean a finite simplicial complex K.

For any two oriented simplexes t_i^r and t_j^{r-1} of consecutive dimensionalities r and $r-1$ of a complex K we define the *incidence coefficient* $(t_i^r : t_j^{r-1})$ as follows:

1°. $(t_i^r : t_j^{r-1}) = 0$, if t_j^{r-1} is not a face of the simplex t_i^r.

2°. Let the vertices of t_i^r be $e_0, \cdots, e_r$ and the vertices of t_j^{r-1} written in the order determining the given oriented simplex t_j^{r-1} be $e_{i_1}, \cdots, e_{i_r}$, where $e_{i_1} \cdots, e_{i_r}$ are all vertices $e_0, \cdots, e_r$ with the exception of one, say e_k. Then

$$(e_k, e_{i_1}, \cdots, e_{i_r}) = \epsilon t_i^r,$$

where, as may be easily seen, the coefficient $\epsilon = \pm 1$ depends only on the oriented simplex t_j^{r-1} (and the oriented simplex t_i^r), but not on the special choice of the order $e_{i_1}, \cdots, e_{i_r}$ determining the oriented simplex t_j^{r-1}. We put

$$(t_i^r : t_j^{r-1}) = \epsilon.$$

It is obvious that

$$(1.10) \qquad (-t_i^r : t_j^{r-1}) = (t_i^r : -t_j^{r-1}) = -(t_i^r : t_j^{r-1}).$$

Further, if $t_i^r = (e_0, \cdots, e_r)$ and $t_j^{r-1} = (e_0, \cdots, e_{k-1}, e_{k+1}, \cdots, e_r)$, then

$$(1.11) \qquad (t_i^r : t_j^{r-1}) = (-1)^k.$$

1.2. The classical construction of combinatorial topology presupposes that a certain commutative group J (generally speaking topological) is given, and is based on consideration of functions $f^r(t^r)$, the argument of which runs through the set of oriented simplexes t^r of the complex K, and whose values are elements of the group J. Besides, it is assumed that always

$$f^r(-t^r) = -f^r(t^r).$$

These functions, called the *r-dimensional functions* on the complex K to the field of coefficients J (also the r-dimensional algebraical complexes or the r-dimensional chains), form in virtue of the operation of addition defined in J a commutative group $L^r(K, J)$ which, if misunderstandings are excluded, we shall denote simply by $L^r(K)$.

The topology defined in J defines a topology also in $L^r(K)$: we obtain a neighbourhood of the element $f_0^r \varepsilon L^r(K)$ by choosing any neighbourhood V of the zero element in J and defining $V(f_0^r)$ as the set of all functions f^r satisfying for all $t^r \varepsilon K$ the condition

$$f^r(t^r) - f_0(t^r) \varepsilon V.$$

1.3. The boundary operators: the lower operator Δ and the upper operator ∇. To each element f^r of the group $L^r(K)$ we correlate: the element Δf^r of the group $L^{r-1}(K)$ and the element ∇f^r of the group $L^{r+1}(K)$ in the following manner:

$$\Delta f^r(t^{r-1}) = \sum_i (t_i^r : t^{r-1}) f^r(t_i^r), \tag{1.30Δ}$$

$$\nabla f^r(t^{r+1}) = \sum_i (t^{r+1} : t_i^r) f^r(t_i^r), \tag{1.30∇}$$

which in virtue of the adopted definition of incidence coefficients means

$$\Delta f^r(e_0, \cdots, e_{r-1}) = \sum_k f^r(e_k, e_0, \cdots, e_{r-1}), \tag{1.31Δ}$$

$$\nabla f^r(e_0, \cdots, e_{r+1}) = \sum_k (-1)^k f^r(e_0, \cdots, \bar{e}_k, \cdots, e_{r+1}), \tag{1.31∇}$$

where, as always in the sequel, $\bar{e}_k$ indicates that e_k must be omitted. We have

$$\Delta\Delta f^r = 0. \tag{1.32Δ}$$

In fact,

$$\Delta\Delta f^r(e_0, \cdots, e_{r-2}) = \sum_k \Delta f^r(e_k, e_0, \cdots, e_{r-2}) = \sum_{h,k} f^r(e_h, e_k, e_0, \cdots, e_{r-2}).$$

But in the last sum for each term $f^r(e_h, e_k, e_0, \cdots, e_{r-2})$ we can find a term

$$f^r(e_k, e_h, e_0, \cdots, e_{r-2}) = -f^r(e_h, e_k, e_0, \cdots, e_{r-2})$$

such that their sum is equal to zero. Also

$$\nabla\nabla f^r = 0. \tag{1.32∇}$$

In fact,

$$\begin{aligned}
\nabla\nabla f^r(e_0, \cdots, e_{r+2}) \\
&= \sum_k (-1)^k \nabla f^r(e_0, \cdots, \bar{e}_k, \cdots, e_{r+2}) \\
&= \sum_k (-1)^k \sum_{h<k} (-1)^h f^r(e_0, \cdots, \bar{e}_h, \cdots, \bar{e}_k, \cdots, e_{r+2}) \\
&\quad + \sum_k (-1)^k \sum_{h>k} (-1)^{h-1} f^r(e_0, \cdots, \bar{e}_k, \cdots, \bar{e}_h, \cdots, e_{r+2}) = 0
\end{aligned}$$

(since interchanging h and k we see that the terms of both sums differ only by the sign).

From the definition of the operators Δ and ∇ it directly follows that they represent homomorphic mappings of $L^r(K, J)$ into $L^{r-1}(K, J)$ and into $L^{r+1}(K, J)$. In the case of a topological group J these homomorphisms are continuous.

DEFINITION 1.33. *If $\Delta f^r = 0$, then the function f^r is called a Δ-cycle; if $\nabla f^r = 0$, then the function f^r is called a ∇-cycle.*

In this terminology the theorems (1.32Δ) and (1.32∇) may be formulated as follows: *the Δ- and ∇-boundaries of every function are respectively Δ- and ∇-cycles.*

DEFINITION 1.34. *The cycle f^r is said to be homologous to zero on K if there exists a function of which f^r is the boundary (of corresponding denotation, i.e. Δ or ∇).*

From the above follows: the r-dimensional Δ-cycles form the kernel of the homomorphism Δ of the group $L^r(K)$ into the group $L^{r-1}(K)$. They form the group $Z^r_\Delta(K, J)$ (or simply $Z^r_\Delta(K)$). Similarly, the r-dimensional ∇-cycles form the group $Z^r_\nabla(K, J)$ or $Z^r_\nabla(K)$—the kernel of the homomorphic mapping ∇ of the group $L^r(K)$ into the group $L^{r+1}(K)$. Hence and from the continuity of the homomorphisms Δ and ∇ it follows that $Z^r_\Delta(K)$ and $Z^r_\nabla(K)$ are closed subgroups of the group $L^r(K)$.

The image of the group $L^r(K)$ under the homomorphism Δ is the group $H^{r-1}(K, J)$ or $H^{r-1}_\Delta(K)$ of all $(r-1)$-dimensional Δ-cycles *homologous to zero in* K. Similarly, the image of the group $L^r(K)$ under the homomorphism ∇ is the group $H^{r+1}_\nabla(K, J)$ or $H^{r+1}_\nabla(K)$ of all $(r+1)$-dimensional ∇-cycles *homologous to zero in* K.

If the group J is bicompact, then so are the groups $L^r(K)$ and also its closed subgroups $Z^r_\Delta(K)$ and $Z^r_\nabla(K)$. The groups $H^r_\Delta(K)$ and $H^r_\nabla(K)$, being images of bicompact groups $L^{r+1}(K)$ and $L^{r-1}(K)$ under continuous homomorphisms Δ and ∇ are bicompact and, consequently, closed in $Z^r_\Delta(K)$ and $Z^r_\nabla(K)$ respectively.

1.4. DEFINITION 1.4. *The factor groups*

$$Z^r_\Delta(K, J) - \overline{H}^r_\Delta(K, J) = B^r_\Delta(K, J),$$

$$Z^r_\nabla(K, J) - \overline{H}^r_\nabla(K, J) = B^r_\nabla(K, J)$$

(where $\overline{H}$ is the closure of the group H in the group Z) are called respectively the r-dimensional Δ- and ∇-Betti groups of the complex K to the field of coefficients J. If J is bicompact or discrete, then obviously $\overline{H} = H$.

Henceforth we shall consider only bicompact and discrete fields of coefficients.

1.5. Let X and Ξ be two commutative groups dual in the sense that each of them is the group of characters of the other (see the Addendum at the end of the paper). Let X be discrete and Ξ bicompact. We consider the functions f^r and ϕ^r belonging respectively to $L^r(K, X)$ and $L^r(K, \Xi)$ and introduce the following notations:

DEFINITION 1.51. (The scalar products of functions.)

$$\phi^r f^r = \sum_{t_i^r \in K} \phi^r(t_i^r) f^r(t_i^r).$$

Correlating to every function $\phi^r \in L^r(K, \Xi)$ the element $\phi^r f^r \in \kappa$, we obtain for a given fixed $f^r \in L^r(K, X)$ a homomorphic mapping of the group $L^r(K, \Xi)$ into κ. In other words, every $f^r \in L^r(K, X)$ may be considered as a character of the group $L^r(K, \Xi)$.

Let us prove that, conversely, every character of the group $L^r(K, \Xi)$ is generated in this sense by an element of the group $L^r(K, X)$.

Let a character h of the group $L^r(K, \Xi)$ be given. Take any simplex $t_i^r \in K$ and consider those functions $\phi^r \in L^r(K, \Xi)$ which only on t_i^r are different from zero. The set of these functions completed by the function identically equal to zero forms a subgroup L_i^r of the group $L^r(K, \Xi)$ isomorphic to the group Ξ. The subgroup L_i^r is mapped by the homomorphism h into κ and this mapping may be considered in virtue of the just established isomorphism between L_i^r and Ξ as an element

$$f_i^h = f^r(t_i^r)$$

of the group X. Thus a function

$$f^r \in L^r(K, X)$$

is defined.

For any $\phi_i^r \in L_i^r$ we have

$$\phi_i^r(t_i^r) f^r(t_i^r) = \phi_i^r(t_i^r) f_i^h = h\phi_i^r \in \kappa.$$

If ϕ^r is an arbitrary element of the group $L^r(K, \Xi)$ and if ϕ_i^r is the function from L_i^r coinciding with ϕ^r on t_i^r and equal to zero on all other simplexes from K, then

$$\phi^r f^r = \sum_i \phi^r(t_i^r) f^r(t_i^r) = \sum_i \phi_i^r(t_i^r) f^r(t_i^r) = \sum_i h\phi_i^r = h \sum_i \phi_i^r = h\phi^r,$$

q.e.d. We have thus proved the following:

THEOREM 1.521. *The group $L^r(K, X)$ is the group of characters of the group $L^r(K, \Xi)$.*

Quite similarly is proved

THEOREM 1.522. *The group $L^r(K, \Xi)$ is the group of characters of the group $L^r(K, X)$.*

THEOREM 1.53. *The homomorphisms Δ and ∇ are conjugate homomor-*

phisms of the group $L^r(K, \Xi)$ into the group $L^{r-1}(K, \Xi)$ and of the group $L^{r-1}(K, X)$ into the group $L^r(K, X)$.

In other words

(1.53) $$\phi^r \nabla f^{r-1} = \Delta \phi^r f^{r-1}.$$

In fact,

$$\begin{aligned}
\overset{r}{\phi}\nabla \overset{r-1}{f} &= \sum_i \overset{r}{\phi}(\overset{r}{t_i}) \sum_j (\overset{r}{t_i} : \overset{r-1}{t_j}) \overset{r-1}{f}(\overset{r-1}{t_j}) = \sum_{i,j} (\overset{r}{t_i} : \overset{r-1}{t_j}) \overset{r}{\phi}(\overset{r}{t_i}) \overset{r-1}{f}(\overset{r-1}{t_j}) \\
&= \sum_j \overset{r-1}{f}(\overset{r-1}{t_j}) \sum_i (\overset{r}{t_i} : \overset{r-1}{t_j}) \phi(\overset{r}{t_i}) = \sum_j \overset{r-1}{f}(\overset{r-1}{t_j}) \Delta \overset{r}{\phi}(\overset{r-1}{t_j}) \\
&= \overset{r-1}{f} \Delta \overset{r}{\phi},
\end{aligned}$$

q.e.d.

1.6. DEFINITION 1.60. *Let M be any subset of the complex K (i.e., a set whose elements are simplexes of the complex K). We shall say that the function $f^r \in L^r(K, J)$ lies on M and write $f^r \subset M$, if for every simplex $t^r \subset K-M$ we have $f^r(t^r)=0$.*

The sum of two functions lying on M evidently lies also on M.

In the sequel M will always be either a subcomplex Q of the complex K or the set $K-Q$ complementary to a subcomplex $Q \subset K$.

DEFINITION 1.61. *The group of all r-dimensional functions $f^r \in L^r(K)$ lying on $K-Q$ will be denoted by $L^r(K-Q)$. The group of all r-dimensional ∇-cycles lying on $K-Q$ will be denoted by $Z^r(K-Q)$.*

Evidently

$$Z_\nabla^r(K-Q) = Z_\nabla^r(K) \cap L^r(K-Q) \subset L^r(K).$$

1.611. *If $f^r \in L^r(K-Q)$, then*

$$\nabla \overset{r}{f} \in Z_\nabla^{r+1}(K-Q).$$

This assertion follows directly from the fact that all faces of a simplex belonging to Q also belong to Q.

DEFINITION 1.612. *A ∇-cycle $f^r \in Z^r(K-Q)$ is said to be homologous to zero on $K-Q$,*

$$f^r \sim 0 \quad on \quad K-Q,$$

if it is the ∇-boundary of a function

$$f^{r-1} \in L^{r-1}(K-Q).$$

The r-dimensional ∇-cycles lying on $K-Q$ and homologous to zero on

$K-Q$ form the subgroup $H^r_\nabla(K-Q)$ of the group $Z^r(K-Q)$ (and of the group $H^r_\nabla(K)$).

The factor group

$$Z^r_\nabla(K-Q) - H^r_\nabla(K-Q)$$

will be denoted by $B^r_\nabla(K-Q, J)$, or simply by $B^r_\nabla(K-Q)$.

DEFINITION 1.62. *Let again Q be a subcomplex of K. The function f^r is said to be a Δ-cycle modulo Q, if $\Delta f^r \subset Q$.*

The r-dimensional Δ-cycles modulo Q evidently form a group; we shall denote it by $Z^r_\Delta(K \bmod Q)$. We have

$$Z^r_\Delta(K \bmod Q) \supset Z^r_\Delta(K).$$

DEFINITION 1.621. *Let f^r be a cycle modulo Q. If there exists a function f^{r+1} such that*

$$\Delta f^{r+1} = f^r + u^r, \qquad u^r \subset Q,$$

then we shall say that f^r is homologous to zero modulo Q on K,

$$f^r \sim 0 \text{ modulo } Q \text{ on } K.$$

The r-dimensional Δ-cycles modulo Q homologous to zero on K modulo Q form a subgroup $H^r_\Delta(K \bmod Q)$ of the group $Z^r_\Delta(K \bmod Q)$. We have

$$H^r_\Delta(K \bmod Q) \supset H^r_\Delta(K).$$

The factor group

$$Z^r_\Delta(K \bmod Q) - H^r_\Delta(K \bmod Q)$$

is denoted by $B^r_\Delta(K \bmod Q)$ (or by $B^r_\Delta(K \bmod Q, J)$).

Let again X and Ξ be two groups dual to each other, X discrete and Ξ bicompact.

THEOREM 1.63. *The annihilator of the group $H^r_\nabla(K-Q, X)$ in the group $L^r(K, \Xi)$ is the group $Z^r_\nabla(K \bmod Q, \Xi)$.*

We have to prove two assertions:

1°. If $\phi^r \in Z^r_\Delta(K \bmod Q, \Xi)$, $f^r \subset H^r_\nabla(K-Q, X)$, then $\phi^r f^r = 0$.

2°. If $\phi^r \in L^r(K, \Xi)$ is a function not belonging to $Z^r_\Delta(K \bmod Q, \Xi)$, then there exists such an $f^r \in H^r_\nabla(K-Q, X)$ that $\phi^r f^r \neq 0$.

Proof of 1°. By assumption

$$f^r = \nabla f^{r-1}, \ f^{r-1} \in L^{r-1}(K-Q, X);$$

then

$$\overset{r}{\phi}\overset{r}{f} = \phi\nabla \overset{r-1}{f} = \Delta\overset{r}{\phi}\overset{r-1}{f} = \sum_i \Delta\overset{r}{\phi}(\overset{r-1}{x_i})\overset{r-1}{f}(\overset{r-1}{x_i}) = 0.$$

Proof of 2°. Suppose that ϕ^r does not enter into $Z^r_\Delta(K \bmod Q)$. Then there exist such a $t_1^{r-1} \varepsilon K-Q$ that

$$\Delta\overset{r}{\phi}(\overset{r-1}{x_1}) = \alpha \neq 0, \qquad \alpha \varepsilon \Xi.$$

Choose an $a \varepsilon X$ such that $\alpha a \neq 0$ and put

$$\overset{r-1}{f}(\overset{r-1}{t_1}) = a,$$
$$\overset{r-1}{f}(\overset{r-1}{t_i}) = 0 \quad \text{for} \quad \overset{r-1}{t_i} \varepsilon K, \quad i \neq 1.$$

Then obviously $f^{r-1} \varepsilon L^{r-1}(K-Q)$. Putting $f^r = \nabla f^{r-1}$, we have

$$\overset{r}{f} \varepsilon H^r_\nabla(K-Q, X),$$
$$\overset{r}{\phi}\overset{r}{f} = \overset{r}{\phi}\nabla\overset{r-1}{f} = \Delta\overset{r}{\phi}\overset{r-1}{f} = \Delta\overset{r}{\phi}(\overset{r}{t_1})\overset{r-1}{f}(\overset{r-1}{t_1}) = \alpha a \neq 0,$$

and Theorem 1.63 is proved.

THEOREM 1.64. *The annihilator of the group* $H^r_\Delta(K \bmod Q, \Xi)$ *in the group* $L^r(K, X)$ *is the group* $Z^r_\nabla(K-Q, X)$.

Proof. We prove, firstly: if

$$\overset{r}{\phi} \varepsilon H^r_\Delta(K \bmod Q, \Xi), \qquad \overset{r}{f} \varepsilon Z^r_\nabla(K-Q, X),$$

then

$$\phi^r f^r = 0.$$

In fact,

$$\phi^r = \Delta\phi^{r+1} + \psi^r, \qquad \psi^r \subset Q,$$
$$\phi^r f^r = \Delta\phi^{r+1}f^r + \psi^r f^r = \phi^{r+1}\nabla f^r = 0.$$

We prove, secondly: if the function $f^r \varepsilon L^r(K, X)$ is not a cycle lying on $K-Q$, then there exists a function $\phi^r \varepsilon H^r_\Delta(K \bmod Q, \Xi)$ such that $\phi^r f^r \neq 0$.

(a) Let f^r be any function not lying on $K-Q$. Then there exists a $t_1^r \varepsilon Q$ such that $f^r(t_1^r) = a \neq 0$. Take such an $\alpha \varepsilon \Xi$ that $\alpha a \neq 0$ and put

$$\overset{r}{\phi}(\overset{r}{t_1}) = \alpha,$$
$$\overset{r}{\phi}(\overset{r}{t_i}) = 0 \quad \text{on other} \quad \overset{r}{t_i} \varepsilon K.$$

Then we have

$$\overset{r}{\phi} \varepsilon H^r_\Delta(K \bmod Q, \Xi),$$
$$\overset{r}{\phi}\overset{r}{f} = \overset{r}{\phi}(\overset{r}{t_1})\overset{r}{f}(\overset{r}{t_1}) = \alpha a \neq 0.$$

(b) Let $f^r \varepsilon L^r(K, X)$ lie on $K-Q$ and not be a ∇-cycle. Then for a certain $t_1^{r+1} \varepsilon K-Q$ we have $\nabla f^r(t_1^{r+1}) = a \neq 0$. Choose $\alpha \varepsilon \Xi$ such that $\alpha a \neq 0$ and put

$$\phi^{r+1}(t_1^{r+1}) = \alpha,$$
$$\phi^{r+1}(t_i^{r+1}) = 0 \quad \text{for other} \quad t_i^{r+1} \varepsilon K.$$

Then

$$\Delta\phi^{r+1} \varepsilon H^r_\Delta(K, \Xi) \subset H^r_\Delta(K \bmod, \Xi),$$

$$\Delta\phi^{r+1} f^r = \phi^{r+1} \nabla f^r = \phi^{r+1}(t_1^{r+1}) \nabla f^r(t_1^{r+1}) = \alpha a \neq 0,$$

and Theorem 1.64 is proved.

Putting in Theorems 1.63 and 1.64

$$Q = 0,$$

we have

1.631. *The annihilator of the group $H^r_\nabla(K, X)$ in the group $L^r(K, \Xi)$ is the group $Z^r_\Delta(K, \Xi)$.*

1.641. *The annihilator of the group $H^r_\Delta(K, \Xi)$ in the group $L^r(K, X)$ is the group $Z^r_\Delta(K, X)$.*

From the theory of characters it is known that if two groups G and Γ are dual to each other and if the subgroup $\Gamma_0 \subset \Gamma$ is the annihilator in Γ of the subgroup $G_0 \subset G$, then, conversely, G_0 is the annihilator in G of the group Γ_0. On ground of this remark we deduce from 1.63 and 1.64

1.632. *The annihilator of the group $Z^r_\Delta(K \bmod Q, \Xi)$ in $L^r(K, X)$ is $H^r_\nabla(K-Q, X)$.*

1.642. *The annihilator of the group $Z^r_\nabla(K-Q, X)$ in $L^r(K, \Xi)$ is $H^r_\Delta(K \bmod Q, \Xi)$.*

1.7. Let A, B, C be three commutative groups, all three discrete or all three bicompact, and let

$$A \supset B, \qquad A - B = C. \tag{1.71}$$

If we denote the groups of characters of the groups A, B, C respectively by A', B', C', then one of the fundamental theorems of the theory of characters may be formulated as follows: *C' is a subgroup of A', namely the annihilator of the group B in the group A', and*

$$B' = A' - C'. \tag{1.72}$$

Or *the annihilator of the subgroup $B \subset C$ in A' is C'.* On ground of this theorem we deduce from 1.63 and 1.64

1.73. *The group* $Z^r_\Delta(K \bmod Q, \Xi)$ *is the group of characters of the group* $L^r(K, X) - H^r_\nabla(K-Q, X)$.

1.74. *The group* $Z^r_\nabla(K-Q, X)$ *is the group of characters of the group* $L^r(K, \Xi) - H^r_\Delta(K \bmod Q, \Xi)$.

In the same way from 1.642 and 1.632 we deduce

1.731. *The group* $H^r_\Delta(K \bmod Q, \Xi)$ *is the group of characters of the group* $L^r(K, X) - Z^r_\nabla(K-Q, X)$.

1.741. *The group* $H^r_\nabla(K-Q, X)$ *is the group of characters of the group* $L^r(K, \Xi) - Z^r_\nabla(K \bmod Q, \Xi)$.

Recall now the so-called second theorem on isomorphisms of Emmy Noether: *If for commutative groups* U, V, W *we have the inclusion relations* $U \supset V \supset W$, *then considering* (in a manner easily understood) $V-W$ *as a subgroup of the group* $U-V$, *we have*

$$(U - W) - (V - W) \approx U - V$$

(where $\approx$ means isomorphic).

Putting

$$U = L^r(K, \Xi), \qquad V = Z^r_\Delta(K \bmod Q, \Xi), \qquad W = H^r_\Delta(K \bmod Q, \Xi),$$

we find

$$[L^r(K, \Xi) - H^r_\Delta(K \bmod Q, \Xi)] - Z^r_\Delta(K \bmod Q, \Xi) - H^r_\Delta(K \bmod Q, \Xi)$$
$$\approx L^r(K, \Xi) - Z^r_\Delta(K \bmod Q, \Xi),$$

wherefrom, on ground of 1.72,

$$\chi[Z^r_\Delta(K \bmod Q, \Xi) - H^r_\Delta(K \bmod Q, \Xi)] = \chi[L^r(K, \Xi) - H^r_\Delta(K \bmod Q, \Xi)]$$
$$- \chi[L^r(K, \Xi) - Z^r_\Delta(K \bmod Q, \Xi)],$$

and on ground of 1.74 and 1.741,

$$\chi[Z^r_\Delta(K \bmod Q, \Xi) - H^r_\Delta(K \bmod Q, \Xi)] \approx Z^r_\Delta(K - Q, X) - H^r_\nabla(K - Q, X),$$

i.e.,

$$\chi[B^r_\Delta(K \bmod Q, \Xi)] = B^r_\nabla(K - Q, X). \tag{1.75}$$

Putting $Q=0$ we obtain as a special case of formula (1.75)

THEOREM 1.751. *The Betti groups* $B^r_\Delta(K, \Xi)$ *and* $B^r_\nabla(K, X)$ *are dual to each other.*

1.8. Consider separately the case $r=0$.

Since a zero-dimensional ∇-cycle cannot be homologous to zero, $B^0_\nabla(K) = Z^0_\nabla(K)$.

Let K be an arbitrary complex. Every function f^0 constant on K (i.e., assuming for all null-dimensional simplexes $t^0 \varepsilon K$ one and the same value) is a ∇-cycle.

In fact, for every one-dimensional simplex $t^1 = (e_0, e_1)$ we have

$$\nabla f^0(t^1) = f^0(e_1) - f^0(e_0) = 0.$$

1.80. *If K is connected and $f^0 \varepsilon Z^0_\nabla(K)$, then f^0 is constant on K.*

In fact, if f^0 is not constant on K, then in virtue of the connectivity of K we may find a $t^1 = (e_0, e_1)$ such that $f^0(e_0) \neq f^0(e_1)$ and, consequently, $\nabla f^0(t^1) \neq 0$.

Thus on connected complexes the zero-dimensional constant functions and only they are cycles.

Hence follows

1.81. *For every complex K the group $Z^0_\nabla(K)$ consists of those and only those zero-dimensional functions which are constant on every component of K.*

1.82. *The group $B^0_\nabla(K, J)$ is a direct sum of groups isomorphic to J and the number of direct summands in this sum is equal to the number of components of the complex K.*

In the group $Z^0_\nabla(K) = B^0_\nabla(K)$ is contained the subgroup $Z^{00}_\nabla(K)$ of those cycles which are constant on the whole K. The factor group

$$Z^0_\nabla(K) - Z^{00}_\nabla(K)$$

we shall denote by $B^{00}_\nabla(K)$.

1.83. *If K consists of p components, then $B^{00}_\nabla(K, J)$ is a direct sum of $p-1$ groups each of which is isomorphic to the group J.*

Defining the group $Z^{00}_\Delta(K)$ as the group of those zero-dimensional functions, the sum of values of which extended over all vertices of K is equal to zero, we have in the group $Z^{00}_\Delta(K)$ a subgroup $H^0_\Delta(K)$ of all zero-dimensional Δ-cycles homologous to zero. The group $B^{00}_\Delta(K) = Z^{00}_\Delta(K) - H^0_\Delta(K)$ is, as is well known, also a direct sum of $p-1$ groups isomorphic to J and, consequently, is isomorphic to the group $B^{00}_\nabla(K)$. The groups $B^{00}_\nabla(K, X)$ and $B^{00}_\Delta(K, \Xi)$ are obviously dual to each other.

2. Simplicial mappings of complexes

2.1. Suppose that to every vertex e^β of a complex K^β corresponds the vertex $Se^\beta = e^\alpha$ of a complex K^α such that to vertices belonging to any simplex of the complex K^β correspond vertices belonging to a simplex of the complex K^α. This correspondence of vertices establishes a mapping S of the complex K^β

into the complex K^α: to every simplex $t_\beta=(e_0^\beta, \cdots, e_r^\beta)$ of the complex K^β corresponds a simplex St_β of the complex K^α with vertices $Se_0^\beta, \cdots, Se_r^\beta$ and the number of dimensions of St_β is less than or equal to the number of dimensions of the simplex t_β. The so-obtained mapping is called a *simplicial mapping of the complex K^β into the complex K^α.*

In virtue of the mapping S, to an *oriented* simplex $t_\beta=(e_0^\beta, \cdots, e_r^\beta)$ of the complex K^β corresponds an oriented simplex (which may be degenerate)

$$t_\alpha = St_\beta = (S\overset{\beta}{e}_0, \cdots, S\overset{\beta}{e}_r)$$

of a complex K^α of the same number of dimensions as t_β.

2.2. A simplicial mapping S of the complex K^β into the complex K^α generates

1°. A homomorphic mapping ρ of the group $L^r(K^\beta)$ into the group $L^r(K^\alpha)$.

2°. A homomorphic mapping σ of the group $L^r(K^\alpha)$ into the group $L^r(K^\beta)$.

Indeed, to every function $f_\beta^r \varepsilon L^r(K^\beta)$ corresponds a function $\rho f_\beta^r \varepsilon L^r(K^\alpha)$ defined by

$$\rho \overset{r}{f}_\beta(\overset{r}{t}_\alpha) = \sum \overset{r}{f}(\overset{r}{t}_\beta), \tag{2.21}$$

where the summation is extended over all $t_\beta^r \varepsilon K^\beta$ such that $St_\beta^r = t_\alpha^r$. To every function $f_\alpha^r \varepsilon L^r(K^\alpha)$ corresponds a function $\sigma f_\alpha^r \varepsilon L^r(K^\beta)$ defined by

$$\sigma \overset{r}{f}_\alpha(\overset{r}{t}_\beta) = \overset{r}{f}(S\overset{r}{t}_\beta). \tag{2.22}$$

THEOREM 2.211. *The homomorphism ρ preserves the lower boundary operator Δ:*

$$\Delta\rho \overset{r}{f}_\beta = \rho\Delta \overset{r}{f}_\beta. \tag{2.211}$$

In fact,

$$\begin{aligned}\Delta\rho \overset{r}{f}_\beta(\overset{\alpha}{e}_1, \cdots, \overset{\alpha}{e}_r) &= \sum_i \rho \overset{r}{f}_\beta(\overset{\alpha}{e}_i, \overset{\alpha}{e}_1, \cdots, \overset{\alpha}{e}_r) \\ &= \sum_i \sum_{j, j_1, \cdots, j_r} \overset{r}{f}_\beta(\overset{\beta}{e}_j, \overset{\beta}{e}_{j_1}, \cdots, \overset{\beta}{e}_{j_r}),\end{aligned}$$

where the summation in the inner sum is extended over all such $j, j_1, \cdots, j_r$ that $Se_j^\beta = e_i^\alpha$, $Se_{j_1}^\beta = e_1^\alpha, \cdots, Se_{j_r}^\beta = e_r^\alpha$. Hence

$$\begin{aligned}\Delta\rho \overset{r}{f}_\beta(\overset{\alpha}{e}_1, \cdots, \overset{\alpha}{e}_r) &= \sum_{j_1, \cdots, j_r} \sum_j \overset{r}{f}_\beta(\overset{\beta}{e}_j, \overset{\beta}{e}_{j_1}, \cdots, \overset{\beta}{e}_{j_r}) \\ &= \sum_{j_1, \cdots, j_r} \Delta \overset{r}{f}_\beta(\overset{\beta}{e}_{j_1}, \cdots, \overset{\beta}{e}_{j_r}) = \rho\Delta \overset{r}{f}_\beta(\overset{\alpha}{e}_1, \cdots, \overset{\alpha}{e}_r).\end{aligned}$$

THEOREM 2.221. *The homomorphism σ preserves the upper boundary operator ∇:*

(2.221) $$\nabla\sigma f^r = \sigma\nabla f^r.$$

In fact, we have the inequalities

$$\begin{aligned}\nabla\sigma f^r_\alpha(\overset{\beta}{e}_0, \cdots, \overset{\beta}{e}_{r+1}) &= \sum_k (-1)^k \sigma f^r_\alpha(\overset{\beta}{e}_0, \cdots, \overset{\beta}{\hat{e}}_k, \cdots, \overset{\beta}{e}_{r+1}) \\ &= \sum_k (-1)^k f^r_\alpha(S\overset{\beta}{e}_0, \cdots, \hat{e}_k, \cdots, S\overset{\beta}{e}_{r+1}) \\ &= \nabla f^r_\alpha(S\overset{\beta}{e}_0, \cdots, S\overset{\beta}{e}_{r+1}) \\ &= \sigma\nabla f^r_\alpha(\overset{\beta}{e}_0, \cdots, \overset{\beta}{e}_{r+1}).\end{aligned}$$

2.3. Let two sub-complexes $Q^\alpha \subset K^\alpha$ and $Q^\beta \subset K^\beta$ be given and let the simplicial mapping S be such that $SQ^\beta \subset Q^\alpha$. Then from

$$f^r_\beta \subset Q^\beta$$

follows

$$\rho f^r_\beta \subset Q^\alpha,$$

and from

$$f^r_\alpha \subset K^\alpha - Q^\alpha$$

follows

$$\sigma f^r_\alpha \subset K^\beta - Q^\beta.$$

Therefore we have the inclusion relations

(2.311) $$\rho Z^r_\Delta(K^\beta \bmod Q^\beta) \subset Z^r_\Delta(K^\alpha \bmod Q^\alpha),$$

(2.312) $$\rho H^r_\Delta(K^\beta \bmod Q^\beta) \subset H^r_\Delta(K^\alpha \bmod Q^\alpha),$$

(2.321) $$\sigma Z^r_\nabla(K^\alpha - Q^\alpha) \subset Z^r_\nabla(K^\beta - Q^\beta),$$

(2.322) $$\sigma H^r_\nabla(K^\alpha - Q^\alpha) \subset H^r_\nabla(K^\beta - Q^\beta).$$

THEOREM 2.33. *The homomorphism ρ generates the homomorphic mapping $\bar{\omega}$ of the group $B^r_\Delta(K^\beta \bmod Q^\beta)$ into the group $B^r_\Delta(K^\alpha \bmod Q^\alpha)$; the homomorphism σ generates the homomorphic mapping π of the group $B^r_\nabla(K^\alpha - Q^\alpha)$ into the group $B^r_\nabla(K^\beta - Q^\beta)$.*

2.4. Let, as always, Ξ and X be two dual groups, Ξ bicompact, X discrete.

THEOREM 2.41. *The homomorphic mapping ρ of the group $L^r(K^\beta, \Xi)$ into the group $L^r(K^\alpha, \Xi)$ and the homomorphic mapping σ of the group $L^r(K^\alpha, \Xi)$ into the group $L^r(K^\beta, X)$ are conjugated.*

In fact, for any $\phi^r_\beta \in L^r(K^\beta, \Xi)$ and $f^r_\alpha \in L^r(K^\alpha, X)$,

$$\phi^r_\beta\sigma f^r_\alpha = \sum_i \phi^r_\beta(t^r_{\beta i}) f^r_\alpha(S t^r_{\beta i}),$$

or, denoting by

$$\sum_{j}^{i}$$

the summation over all $t^r_{\beta j}$ such that $St^r_{\beta j}=t^r_{\alpha i}$ (and zero, if such $t^r_{\beta j}$ do not exist),

$$\begin{aligned}\phi^r_\beta \sigma f^r_\alpha &= \sum_i \left[\sum_j^i \phi^r_\beta(t^r_{\beta j}) \right] f^r_\alpha(t^r_{\alpha i}) \\ &= \sum_i \rho f^r_\beta(t^r_{\alpha i}) f^r_\alpha(t^r_{\alpha i}) = \rho f^r_\beta \cdot f^r_\alpha.\end{aligned}$$

THEOREM 2.42. *The homomorphic mapping $\bar{\omega}$ of the group $B^r_\Delta(K^\beta \bmod Q^\beta, \Xi)$ into the group $B^r_\Delta(K^\alpha \bmod Q^\alpha, \Xi)$ and the homomorphic mapping π of the group $B^r_\nabla(K^\beta - Q^\beta, X)$ into the group $B^r_\nabla(K^\alpha - Q^\alpha, X)$ are conjugated.*

In fact, let $\zeta^\beta \varepsilon B^r_\Delta(K^\beta \bmod Q^\beta, \Xi)$, $z^\alpha \varepsilon B^r_\nabla(K^\alpha - Q^\alpha, X)$ be chosen arbitrarily. Choose the cycles $\phi^r_\beta \varepsilon \zeta^\beta$ and $f^r_\alpha \varepsilon z^\alpha$. Then $\bar{\omega}\zeta^\beta$ is an element of the group $B^r_\Delta(K^\alpha \bmod Q^\alpha, \Xi)$ containing $\rho\phi^r_\beta$ and πz^α is an element of the group $B^r_\nabla(K^\beta - Q^\beta, X)$ containing σf^r_α. We have

$$\zeta^\beta \cdot \pi z^\alpha = \phi^r_\beta \cdot \sigma f^r_\alpha = \rho \phi^r_\beta \cdot f^r_\alpha = \bar{\omega}\zeta^\beta \cdot z^\alpha,$$

q.e.d. From the result just proved directly follows

2.43. *Let S and S' be two simplicial mappings of the complex K^β into the complex K^α such that for given subcomplexes $Q^\beta \subset K^\beta$ and $Q^\alpha \subset K^\alpha$ we have $S(Q^\beta) \subset Q^\alpha$, $S'(Q^\beta) \subset Q^\alpha$. Let both mappings S and S' generate one and the same homomorphism $\bar{\omega}$ of the group $B^r_\Delta(K^\beta \bmod Q^\beta, \Xi)$ into the group $B^r_\Delta(K^\alpha \bmod Q^\alpha, \Xi)$. Then the mappings S and S' generate one and the same homomorphism π of the group $B^r_\nabla(K^\alpha - Q^\alpha, X)$ into the group $B^r_\nabla(K^\beta - Q^\beta, X)$.*

The dual formulation is, of course, also true.

The following remark is essential for the sequel. Let the simplicial mappings S_0 and S_1 of the complex K^β into the complex K^α satisfy the condition

2.44. *Whatever be the simplex $t_\beta \subset K^\beta$ there exists a simplex $t_\alpha \subset K^\alpha$ having among its faces the simplexes $S_0 t_\beta$ and $S_1 t_\beta$ and if $t_\beta \subset Q^\beta$, then we may suppose that $t_\alpha \subset Q^\alpha$.*

In this case the mappings S_0 and S_1 are evidently homotopic and if we denote by S_u, $0 \leqq u \leqq 1$, the deformation of S_0 into S_1, we may suppose that for any u we have $S_u \bar{Q}^\beta \subset \bar{Q}^\alpha$. Hence it follows that S_0 and S_1 generate one and the same homomorphism $\bar{\omega}$ of the group $B^r_\Delta(K^\beta \bmod Q^\beta, \Xi)$ into the group $B^r_\Delta(K^\alpha \bmod Q^\alpha, \Xi)$. From 2.43 it follows that S_0 and S_1 generate one and the same homomorphism π of the group $B^r_\nabla(K^\alpha - Q^\alpha, X)$ into the group $B^r_\nabla(K^\beta - Q^\beta, X)$.

2.45. *If the simplicial mappings S_0 and S_1 satisfy the condition* 2.44, *then*

they generate one and the same homomorphism $\bar{\omega}$ of the group $B^r_\Delta(K^\beta \bmod Q^\beta, \Xi)$ into the group $B^r_\Delta(K^\alpha \bmod Q^\alpha, \Xi)$ and one and the same homomorphism π of the group $B^r_\nabla(K^\alpha - Q^\alpha, X)$ into the group $B^r_\nabla(K^\beta - Q^\beta, X)$(7).

3. Spectra and their limit groups

3.1. A partially ordered set D shall be called ***unbounded***, if for any two elements d_1 and d_2 of the set D there exists an element d_3 following after d_1 as well as after d_2:

$$d_3 > d_1, \quad d_3 > d_2.$$

Consider the unbounded partially ordered set Δ consisting of the groups H^α; for the sake of simplicity we shall suppose that the indices α are ordinal numbers (the order of which, however, must not be at all connected with the order in Δ). Suppose that for any two groups H^α and H^β which are elements of Δ and satisfy in Δ the condition $H^\beta \geqq H^\alpha$ a homomorphic mapping $\bar{\omega}^\beta_\alpha$ (*projection*) of the group H^β into the group H^α is defined such that for $H^\gamma \geqq H^\beta \geqq H^\alpha$ we have always

$$\bar{\omega}^\beta_\alpha \bar{\omega}^\gamma_\beta = \bar{\omega}^\gamma_\alpha$$

and $\bar{\omega}^\beta_\alpha$ is the identical mapping.

The system of groups H^α and mappings $\bar{\omega}^\beta_\alpha$ is called the ***inverse spectrum*** and is denoted by $[H^\alpha; \bar{\omega}^\beta_\alpha]$.

Every inverse spectrum defines the limit group

$$H = \varprojlim\, [H^\alpha; \bar{\omega}^\beta_\alpha].$$

The elements of the group H are the *threads* of the spectrum $[H^\alpha; \bar{\omega}^\beta_\alpha]$, i.e., systems of elements $\eta = \{\eta^\alpha\}$ satisfying the following conditions:

1°. Every η^α is an element of the group H^α and η contains only one element of each group.

2°. If η^α and η^β are elements of the thread η and $\eta^\alpha \in H^\alpha$, $\eta^\beta \in H^\beta$, then $\bar{\omega}^\beta_\alpha \eta^\beta = \eta^\alpha$.

If $\eta_1 = \{\eta_1^\alpha\}$ and $\eta_2 = \{\eta_2^\alpha\}$ are two threads of the inverse spectrum $[H^\alpha; \bar{\omega}^\beta_\alpha]$, then $\eta = \{\eta_1^\alpha + \eta_2^\alpha\}$ is also a thread and we put

$$\eta = \eta_1 + \eta_2.$$

Fixing α and correlating to every thread η the element η^α contained in it, we obtain a homomorphic mapping $\bar{\omega}_\alpha$ of the group H into the group H^α. The inverse spectra are considered always under the assumption that the groups H^α are bicompact. Then the group $H = \varprojlim [H^\alpha; \bar{\omega}^\beta_\alpha]$ will be also topologized;

(7) A purely combinatorial proof of this theorem has been given by Čech, *Théorie générale de l'homologie*, Fundamenta Mathematicae, vol. 19 (1932), pp. 149–183, especially pp. 158–159 (§12).

a neighbourhood of the thread $\eta_0=\{\eta_0^\alpha\}$ is obtained, if we choose a finite number $\eta_0^{\alpha_1}, \cdots, \eta_0^{\alpha_s}$ of its elements, choose for each of them a neighbourhood $O\eta_0^{\alpha_i}$ in H^{α_i} and take all threads $\eta=\{\eta^\alpha\}$ satisfying the conditions $\eta^{\alpha_i} \varepsilon O\eta_0^{\alpha_i}$ for $i=1, 2, \cdots, s$. The so-topologized group H proves to be bicompact.

3.2. Suppose that we have an unbounded partially ordered set D of discrete groups $\mathfrak{H}^\alpha$ and assume, further, for convenience that all elements h^α of the group $\mathfrak{H}^\alpha$ are different from the elements h^β of the group $\mathfrak{H}^\beta$, if $\alpha\neq\beta$.

Suppose further that for any two groups $\mathfrak{H}^\alpha$, $\mathfrak{H}^\beta$ such that $\mathfrak{H}^\alpha<\mathfrak{H}^\beta$ in D is established a homomorphic mapping π_β^α (projection) of the group $\mathfrak{H}^\alpha$ into the group $\mathfrak{H}^\beta$ such that for $\mathfrak{H}^\alpha\leqq\mathfrak{H}^\beta\leqq\mathfrak{H}^\gamma$ we have $\pi_\gamma^\beta\pi_\beta^\alpha=\pi_\gamma^\alpha$ and π_α^α is the identical mapping. The system of groups $\mathfrak{H}^\alpha$ and homomorphisms π_β^α is called the *direct spectrum* $[\mathfrak{H}^\alpha; \pi_\beta^\alpha]$. The set-theoretical sum $\bigcup_\alpha\mathfrak{H}^\alpha$ of all groups $\mathfrak{H}^\alpha$ is called the *spectral set* of the given direct spectrum. Two elements of the spectral set, $h^\alpha \varepsilon \mathfrak{H}^\alpha$ and $h^\beta \varepsilon \mathfrak{H}^\beta$ are called equivalent if there exists in the spectrum a group $\mathfrak{H}^\gamma$ such that $\mathfrak{H}^\gamma>\mathfrak{H}^\alpha$, $\mathfrak{H}^\gamma>\mathfrak{H}^\beta$ and $\pi_\gamma^\alpha h^\alpha=\pi_\gamma^\beta h^\beta$.

This notion of equivalence obviously possesses the properties of reflextivity, symmetry and, in virtue of the unboundedness of the partially ordered set D, also of transitivity. The spectral set $\bigcup_\alpha\mathfrak{H}^\alpha$ falls therefore into classes of equivalence which we shall for the sake of shortness call the *bundles* of the direct spectrum $[\mathfrak{H}^\alpha; \pi_\beta^\alpha]$. The bundles possess the following obvious property: *every projection of an element of any bundle is an element of the same bundle.* Hence follows: If $h^\alpha \varepsilon \mathfrak{H}^\alpha$ is an element of the bundle h and $\mathfrak{H}^\beta>\mathfrak{H}^\alpha$, then in $\mathfrak{H}^\beta$ there is an element h^β of the bundle h.

In any two bundles h_1 and h_2 we may find elements h_1^α and h_2^α belonging to one and the same $\mathfrak{H}^\alpha$. In fact, choose arbitrarily $h^{\alpha_1} \varepsilon h_1$ and $h^{\alpha_2} \varepsilon h_2$ and take $\mathfrak{H}^\alpha>\mathfrak{H}^{\alpha_1}$, $\mathfrak{H}^\alpha>\mathfrak{H}^{\alpha_2}$. Then

$$\pi_\alpha^{\alpha_1} h^{\alpha_1} = h_1^\alpha \varepsilon h_1, \qquad \pi_\alpha^{\alpha_2} h^{\alpha_2} = h_2^\alpha \varepsilon h_2.$$

Let h_1 and h_2 be two bundles. From the above follows that we can find two elements $h_1^\alpha \varepsilon h_1$, $h_2^\alpha \varepsilon h_2$ belonging to one and the same group $\mathfrak{H}^\alpha$. We shall call the bundle h, containing the element $h^\alpha=h_1^\alpha+h_2^\alpha$ the sum of the bundles h_1 and h_2. This definition does not depend on the choice of the elements $h_1^\alpha \varepsilon h_1$ and $h_2^\alpha \varepsilon h_2$. In fact, if $h_1^\beta \varepsilon h_1$, $h_2^\beta \varepsilon h_2$, then $h^\beta=h_1^\beta+h_2^\beta$ belongs to the same bundle as h^α. In order to prove this observe that since h_1^α and h_1^β belong to one and the same bundle, there exists such an $\mathfrak{H}^{\gamma_1}$,

$$\mathfrak{H}^{\gamma_1} > \mathfrak{H}^\alpha, \ \mathfrak{H}^{\gamma_1} > \mathfrak{H}^\beta,$$

that

$$\pi_{\gamma_1}^\alpha h_1^\alpha = \pi_{\gamma_1}^\beta h_1^\beta = h^{\gamma_1}.$$

Similarly there exists an $\mathfrak{H}^{\gamma_2}$ such that

$$\pi^{\alpha}_{\gamma_2} h^{\alpha}_2 = \pi^{\beta}_{\gamma_2} h^{\beta}_2 = h^{\gamma_2}.$$

Take an $\mathfrak{H}^{\gamma}$ following after $\mathfrak{H}^{\gamma_1}$ as well as after $\mathfrak{H}^{\gamma_2}$. Then

$$\pi^{\gamma_1}_{\gamma} h^{\gamma}_1 = \pi^{\alpha}_{\gamma} h^{\alpha}_1 = \pi^{\beta}_{\gamma} h^{\beta}_1 = h^{\gamma}_1; \qquad \pi^{\gamma_2}_{\gamma} h^{\gamma}_2 = \pi^{\alpha}_{\gamma} h^{\alpha}_2 = \pi^{\beta}_{\gamma} h^{\beta}_2 = h^{\gamma}_2$$

and, consequently,

$$h^{\gamma} = h^{\gamma}_1 + h^{\gamma}_2 = \pi^{\alpha}_{\gamma} h^{\alpha} = \pi^{\beta}_{\gamma} h^{\beta},$$

so that the equivalence of $h^{\alpha} = h^{\alpha}_1 + h^{\alpha}_2$ and $h^{\beta} = h^{\beta}_1 + h^{\beta}_2$ is proved.

The so-defined addition of bundles is obviously associative. It turns the set of all bundles into a (discrete) group $\mathfrak{H} = \lim_{\rightarrow} [\mathfrak{H}^{\alpha}; \pi^{\alpha}_{\beta}]$ which is called the *limit group of the direct spectrum* $[\mathfrak{H}^{\alpha}; \pi^{\alpha}_{\beta}]$. The zero elements of all groups $\mathfrak{H}^{\alpha}$ are all equivalent to each other and, consequently, belong to one and the same bundle—the zero bundle, which is the zero element of the group $\mathfrak{H}$.

The elements $-h^{\alpha}$ opposite to the elements h^{α} of a bundle h form the bundle $-h$.

Correlating to every element h^{α} of the group $\mathfrak{H}^{\alpha}$ the bundle containing this element, we obtain a homomorphic mapping π^{α} of the group $\mathfrak{H}^{\alpha}$ into the group $\mathfrak{H}$. For $\mathfrak{H}^{\alpha} < \mathfrak{H}^{\beta}$ we have

$$\pi^{\alpha} = \pi^{\beta} \pi^{\alpha}_{\beta}.$$

3.3. Let two spectra—the direct spectrum

$$[H^{\alpha}; \bar{\omega}^{\beta}_{\alpha}]$$

and the inverse spectrum

$$[\mathfrak{H}^{\alpha}; \pi^{\alpha}_{\beta}]$$

—be given. If the groups H^{α} and $\mathfrak{H}^{\alpha}$ composing these spectra are dual to each other for every given α and the homomorphisms $\bar{\omega}^{\beta}_{\alpha}$ and π^{α}_{β} are conjugate, then the spectra are said to be conjugate to each other. It is known that the limit groups H and $\mathfrak{H}$ of two conjugate spectra are dual (Steenrod(4)) and that an element h of the group $\mathfrak{H} = \lim_{\rightarrow} [\mathfrak{H}^{\alpha}; \pi^{\alpha}_{\beta}]$ realizes a homomorphism of the group $H = \lim_{\leftarrow} [H^{\alpha}; \bar{\omega}^{\beta}_{\alpha}]$ in κ according to the formula:

$$\eta^{h} = \eta^{\alpha} h^{\alpha},$$

where η^{α} and h^{α} are taken arbitrarily in η and h: it turns out that the so-defined homomorphism does not depend on the elements of arbitrariness involved in its definition.

3.4. In the sequel we shall almost exclusively consider direct spectra; therefore by a "spectrum" without any adjective we shall understand a direct spectrum and in accordance with this omit the arrows in the formulae.

3.41. *Let two spectra*

$$[\overset{\alpha}{U};\ \overset{\alpha}{\pi_\beta}], \qquad [\overset{\alpha}{V};\ \overset{\alpha}{\rho_\beta}]$$

be given, the elements of which correspond to each other in one-to-one manner ($U^\alpha \rightleftarrows V^\alpha$). *For every* α *let there be given a homomorphic mapping* ϕ_α *of the group* U^α *in the group* V^α. *If for every* α, *every* U^α, *and every* $u^\alpha \varepsilon U^\alpha$ *the condition*

$$\overset{\alpha}{\rho_\beta}\phi_\alpha \overset{\alpha}{u} = \phi_\beta \overset{\alpha}{\pi_\beta}\overset{\alpha}{u}$$

is satisfied, then we obtain a homomorphic mapping ϕ *of the group* $U = \lim\ [U^\alpha;\ \pi_\beta^\alpha]$ *in the group* $V = \lim\ [V^\alpha;\ \rho_\beta^\alpha]$ *as the mapping correlating to every bundle* u *of the spectrum* $[U^\alpha;\ \pi_\beta^\alpha]$ *the bundle* $v = \phi u$ *of the spectrum* $[V^\alpha;\ \rho_\beta^\alpha]$ *containing with an element* $u^\alpha \varepsilon U^\alpha$ *also the element* $\phi_\alpha u^\alpha$.

It suffices to show: from $u^\alpha \equiv u^\beta$ (where $\equiv$ is the sign of equivalence) follows $\phi_\alpha u^\alpha \equiv \phi_\beta u^\beta$. But if $u^\alpha \equiv u^\beta$, then there exists such a U^γ that

$$\overset{\alpha}{\pi_\gamma}\overset{\alpha}{u} = \overset{\beta}{\pi_\gamma}\overset{\beta}{u}$$

and, consequently,

$$\overset{\alpha}{\rho_\gamma}\phi_\alpha \overset{\alpha}{u} = \phi_\gamma \overset{\alpha}{\pi_\gamma}\overset{\alpha}{u} = \phi_\gamma \overset{\beta}{\pi_\gamma}\overset{\beta}{u} = \overset{\beta}{\rho_\gamma}\phi_\beta \overset{\beta}{u},$$

q.e.d.

We add two remarks, the proofs of which may be left to the reader.

3.411. *If, whatever be* α *and* $v^\alpha \varepsilon V^\alpha$, *there is a* $V^\beta > V^\alpha$ *and a* $u^\beta \varepsilon U^\beta$ *such that*

$$\phi_\beta \overset{\beta}{u} = \overset{\alpha}{\rho_\beta}\overset{\alpha}{v},$$

then the mapping ϕ *is a homomorphism of the group* U *on the group* V.

3.412. *If from* $\phi u = 0$ *it follows that* $u \varepsilon U$ *contains the zero element of some group* U^α, *then* ϕ *is an isomorphic mapping.*

3.5. A partially ordered set D is called a *part* of the partially ordered set D', if every element of D is an element of D' and if from $d_1 > d_2$ in D follows $d_1 > d_2$ in D' (but it is not demanded that from $d_1 > d_2$ in D', $d_1 \varepsilon D$, $d_2 \varepsilon D$, should necessarily follow $d_1 > d_2$ in D).

A part D of an unbounded partially ordered set D' is called *cofinal* to the whole D' if D is unbounded and after every element of D' in D' follows an element of D.

A spectrum I is called a part, respectively a cofinal part, of a spectrum II, if the spectrum I considered as a partially ordered set of groups, of which it is composed, is a part, respectively a cofinal part, of the spectrum II and if the projections in I coincide with corresponding projections in II.

Let now the spectrum I form a part of the spectrum II. It is obvious that two elements of the spectral set of I equivalent in the spectrum I are equiva-

lent also in the spectrum II. If I is a cofinal part of II, then, conversely, two elements of I equivalent in the spectrum II are equivalent also in the spectrum I. From the first assertion it follows that every bundle of the spectrum I is contained in a bundle of the spectrum II. From the second assertion it follows that in the case of cofinality of I and II every bundle of the spectrum II contains only one bundle of the spectrum I (that every bundle of II contains at least one bundle of I follows directly from the condition of cofinality). Thus we have

3.51. *If one of two spectra forms a cofinal part of the other, then both spectra have isomorphic limit groups.*

3.6. We shall also need the following proposition.

3.61. *Suppose that the spectra*

$$[U^{\alpha}; \pi^{\alpha}_{\beta}], \qquad [U^{\alpha\lambda}; \pi^{\alpha\lambda}_{\beta\mu}]$$

satisfy the following conditions:

1°. *There exists an isomorphic mapping* $\phi^{\alpha\lambda}_{\alpha}$ *of the group* $U^{\alpha\lambda}$ *on the group* U^{α}.

2°. *If* $U^{\beta\mu} > U^{\alpha\lambda}$, *then* $U^{\beta} > U^{\alpha}$.

3°. *For every* $u^{\alpha\lambda} \varepsilon U^{\alpha\lambda}$ *we have*

$$\pi^{\alpha\lambda}_{\beta\mu} u^{\alpha\lambda} = (\phi^{\beta\mu}_{\mu})^{-1} \pi^{\alpha}_{\beta} \phi^{\alpha\lambda}_{\alpha} u^{\alpha\lambda}.$$

Then the limit groups lim $[U^{\alpha\lambda}; \pi^{\alpha\lambda}_{\beta\mu}]$ *and* lim $[U^{\alpha}; \pi^{\alpha}_{\beta}]$ *are isomorphic.*

For the proof construct first from the groups $U^{\alpha\lambda}$ a partially ordered set D' by putting *always* $U^{\beta\mu} > U^{\alpha\lambda}$, if in the spectrum $[U^{\alpha}; \pi^{\alpha}_{\beta}]$ the inequality $U^{\beta} > U^{\alpha}$ holds. The partially ordered set D' is evidently unbounded.

We define now the projections

$$\rho^{\alpha\lambda}_{\beta\mu} u^{\alpha\lambda} = (\phi^{\beta\mu}_{\beta})^{-1} \pi^{\alpha}_{\beta} \phi^{\alpha\lambda}_{\alpha} u^{\alpha\lambda}.$$

Since

$$\rho^{\beta\mu}_{\gamma\nu} \rho^{\alpha\lambda}_{\beta\mu} = (\phi^{\gamma\nu}_{\gamma})^{-1} \pi^{\beta}_{\gamma} \phi^{\beta\mu}_{\beta} (\phi^{\beta\mu}_{\beta})^{-1} \pi^{\alpha}_{\beta} \phi^{\alpha\lambda}_{\alpha} = (\phi^{\gamma\nu}_{\gamma})^{-1} \pi^{\alpha}_{\gamma} \phi^{\alpha\lambda}_{\alpha} = \rho^{\alpha\lambda}_{\gamma\nu},$$

$[U^{\alpha\lambda}; \rho^{\alpha\lambda}_{\beta\mu}]$ is obviously a spectrum containing the spectrum $[U^{\alpha\lambda}; \pi^{\alpha\lambda}_{\beta\mu}]$ as its cofinal part. Therefore the groups lim $[U^{\alpha\lambda}; \rho^{\alpha\lambda}_{\beta\mu}]$ and lim $[U^{\alpha\lambda}; \pi^{\alpha\lambda}_{\beta\mu}]$ are isomorphic. Identifying for every α the group U^{α} with the group $U^{\alpha 1}$ isomorphic to it, we may also consider the spectrum $[U^{\alpha}; \pi^{\alpha}_{\beta}]$ as a part of the spectrum $[U^{\alpha\lambda}; \rho^{\alpha\lambda}_{\beta\mu}]$ and even as a cofinal part. In fact, in order to obtain $U^{\beta 1}$ following after a given $U^{\alpha\lambda}$ in the spectrum $[U^{\alpha\lambda}; \rho^{\alpha\lambda}_{\beta\mu}]$ it is sufficient to take $U^{\beta} > U^{\alpha}$ in the spectrum $[U^{\alpha}; \pi^{\alpha}_{\beta}]$; evidently, $U^{\beta 1} > U^{\alpha\lambda}$ in the spectrum $[U^{\alpha\lambda}; \rho^{\alpha\lambda}_{\beta\mu}]$.

Thus the group lim $[U^{\alpha}; \pi^{\alpha}_{\beta}]$ is isomorphic to the group lim $[U^{\alpha\lambda}; \rho^{\alpha\lambda}_{\beta\mu}]$ and, consequently, also to the group lim $[U^{\alpha\lambda}; \pi^{\alpha\lambda}_{\beta\mu}]$, q.e.d.

4. The formal duality

4.1. Let there be given an unbounded partially ordered set $\mathfrak{K}$ of complexes K^α. Suppose that in every complex K^α are given two subcomplexes K_0^α and C^α. The complex C^α is called a special subcomplex of the complex K^α. If $K^\beta > K^\alpha$, then simplicial mappings S_α^β of the complex K^β into the complex K^α are defined, which are called *projections* of K^β into K^α; they satisfy the following conditions:

1°. For any projection S_α^β

$$S_\alpha^\beta(K_0^\beta) \subset K_0^\alpha, \qquad S_\alpha^\beta(C^\beta) \subset C^\alpha.$$

2°. If S_α^β and $\bar{S}_\alpha^\beta$ are two projections of K^β into K^α, then for any simplex $t_\beta \varepsilon K^\beta$ the simplexes $S_\alpha^\beta t_\beta$ and $\bar{S}_\alpha^\beta t_\beta$ are faces of a certain simplex $t_\alpha \varepsilon K^\alpha$ and, if $t_\beta \varepsilon K_0^\beta$ or $t_\beta \varepsilon C^\beta$, then we may assume that correspondingly $t_\alpha \varepsilon K_0^\alpha$, $t_\alpha \varepsilon C^\alpha$.

From the condition 1° it follows that every projection S_α^β generates a homomorphism σ_β^α of the groups

$$L^r(K^\alpha - C^\alpha), \qquad L^r(K_0^\alpha - C^\alpha), \qquad L^r(K^\alpha - K_0^\alpha - C^\alpha)\text{(8)}$$

respectively into the groups

$$L^r(K^\beta - C^\beta), \qquad L^r(K_0^\beta - C^\beta), \qquad L^r(K^\beta - K_0^\beta - C^\beta);$$

the homomorphisms σ_β^α generate further homomorphical mappings π_β^α of the groups

$$B_\nabla^r(K_0^\alpha - C^\alpha), \qquad B_\nabla^r(K^\alpha - K_0^\alpha - C^\alpha) \tag{4.11}$$

into the groups

$$B_\nabla^r(K_0^\beta - C^\beta), \qquad B_\nabla^r(K^\beta - K_0^\beta - C^\beta) \tag{4.12}$$

and from the condition 2° it follows that *all projections of the complex K^β into the complex K^α generate one and the same homomorphism π_β^α of the groups* (4.11) *into the groups* (4.12). These homomorphisms are also called projections.

Suppose that beside the conditions 1° and 2° the following condition is also satisfied:

3°. If $K^\gamma > K^\beta > K^\alpha$, then whatever be the projections S_α^β and S_β^γ of respectively K^β into K^α and K^γ into K^β, the simplicial mapping $S_\alpha^\beta S_\beta^\gamma$ of the complex K^γ into the complex K^α is a projection.

From the condition 3° it follows that

$$\pi_\gamma^\beta \pi_\beta^\alpha = \pi_\gamma^\alpha$$

and that we have the spectra

(8) We recall that $K_0^\alpha - C^\alpha$ means the set of all simplexes of K_0^α which do not belong to C, i.e., $K_0^\alpha - C^\alpha = K_0^\alpha - K_0^\alpha \cap C^\alpha$.

$$[B^r_\nabla(K^\alpha - C^\alpha); \pi^\alpha_\beta], \qquad [B^r_\nabla(K^\alpha_0 - C^\alpha); \pi^\alpha_\beta],$$
$$[B^r_\nabla(K^\alpha - K^\alpha_0 - C^\alpha); \pi^\alpha_\beta].$$

DEFINITION 4.10. *A partially ordered set $\mathfrak{R}$ of complexes K^α satisfying the conditions 1°, 2° and 3° is called simply-connected with respect to the dimensionality r for given special subcomplexes C^α, if the group*

$$\lim\,[B^r_\nabla(K^\alpha - C^\alpha); \pi^\alpha_\beta]$$

consists only of the zero element.

The aim of the present paragraph is the proof of the following

THEOREM 4.1. *If $\mathfrak{R}$ is simply connected with respect to the dimensionalities r and $r+1$, then the groups*

$$\lim\,[B^r_\nabla(K^\alpha_0 - C^\alpha); \pi^\alpha_\beta], \qquad \lim\,[B^{r+1}_\nabla(K^\alpha - K^\alpha_0 - C^\alpha); \pi^\alpha_\beta]$$

are isomorphic.

4.2. **Preliminary remarks to the proof of Theorem** 4.1. For the sake of shortness we shall write
L^r_α instead of $L^r(K^\alpha - C^\alpha)$,
$L^r_{0\alpha}$ instead of $L^r(K^\alpha_0 - C^\alpha)$,
$L^r_{g\alpha}$ instead of $L^r(K^\alpha - K^\alpha_0 - C^\alpha)$,
$Z^r_{0\alpha}$ instead of $Z^r_\nabla(K^\alpha_0 - C^\alpha)$,
$Z^r_{g\alpha}$ instead of $Z^r_\nabla(K^\alpha - K^\alpha_0 - C^\alpha)$,
$H^r_{0\alpha}$ instead of $H^r_\nabla(K^\alpha_0 - C^\alpha)$,
$H^r_{g\alpha}$ instead of $H^r_\nabla(K^\alpha - K^\alpha_0 - C^\alpha)$,
$B^r_{0\alpha}$ instead of $B^r_\nabla(K^\alpha_0 - C^\alpha)$,
$B^r_{g\alpha}$ instead of $B^r_\nabla(K^\alpha - K^\alpha_0 - C^\alpha)$,
B^r_0 instead of $\lim\,[B^r_\nabla(K^\alpha_0 - C^\alpha); \pi^\alpha_\beta]$,
B^r_g instead of $\lim\,[B^r_\nabla(K^\alpha - K^\alpha_0 - C^\alpha); \pi^\alpha_\beta]$.
The elements of the groups

$$B^r_{0\alpha}, \qquad B^r_0, \qquad B^{r+1}_{g\alpha}, \qquad B^{r+1}_g$$

we shall denote respectively by

$$\overset{r}{u}_\alpha, \qquad \overset{r}{u}, \qquad \overset{r+1}{v}_\alpha, \qquad \overset{r+1}{v}.$$

The elements of the groups

$$L^r_\alpha, \qquad L^r_{0\alpha}, \qquad L^r_{g\alpha},$$

we shall denote respectively by

$$\overset{r}{f}_\alpha, \qquad \overset{r}{f}_{0\alpha}, \qquad \overset{r}{g}_\alpha.$$

If some function f_α^r is given, we denote by $\bar{f}_\alpha^r$ the function which is equal to f_α^r on K_0^α and is equal to zero on $K^\alpha - K_0^\alpha$; by Af_α^r we denote the function $f_{0\alpha}^r$ equal to f_α^r on K_0^α.

If a function $f_{0\alpha}^r$ is given, we denote by $Ef_{0\alpha}^r$ the function f_α^r equal to $f_{0\alpha}^r$ on K_0^α and to zero on $K^\alpha - K_0^\alpha$.

Let us now formulate some simple properties of the operators A and E:

$$EAf_\alpha^r = \bar{f}_\alpha^r, \tag{4.21}$$

$$AEf_{0\alpha}^r = f_{0\alpha}^r. \tag{4.22}$$

From

$$g_\alpha^r \in L_{g\alpha}^r$$

follows

$$A_{g\alpha}^r = 0. \tag{4.23}$$

If $t_{0\alpha}^{r+1} \in K_0^\alpha$, then

$$\nabla Af_\alpha^r(t_{0\alpha}^{r+1}) = \nabla f_\alpha^r(t_{0\alpha}^{r+1}), \tag{4.24}$$

$$\nabla Ef_{0\alpha}^r(t_{0\alpha}^{r+1}) = \nabla f_{0\alpha}^r(t_{0\alpha}^{r+1}). \tag{4.25}$$

Further,

$$\nabla Af_\alpha^r = A\nabla f_\alpha^r. \tag{4.26}$$

In fact, if $t_{0\alpha}^{r+1} \in K_0^\alpha$, then $\nabla Af_\alpha^r(t_{0\alpha}^{r+1})$ and $A\nabla f_\alpha^r(t_{0\alpha}^{r+1})$ coincide with $\nabla f_\alpha^r(t_{0\alpha}^{r+1})$.

If S_α^β is any projection of K^β into K^α, then

$$A\overset{\alpha}{\sigma}_\beta f_\alpha^r = \overset{\alpha}{\sigma}_\beta Af_\alpha^r. \tag{4.27}$$

In fact, from $t_{0\beta}^r \in K_0^\beta$ follows

$$A\overset{\alpha}{\sigma}_\beta f_\alpha^r(t_{0\beta}^r) = \overset{\alpha}{\sigma}_\beta f_\alpha^r(t_{0\beta}^r) = f_\alpha^r(\overset{\beta}{S}_\alpha t_{0\beta}^r), \qquad \overset{\alpha}{\sigma}_\beta Af_\alpha^r(t_{0\beta}^r) = Af_\alpha^r(\overset{\beta}{S}_\alpha t_{0\beta}^r) = f_\alpha^r(\overset{\beta}{S}_\alpha t_{0\beta}^r).$$

Observe, finally, that from $f_{0\alpha}^r \in Z_{0\alpha}^r$ follows $\nabla Ef_{0\alpha}^r \in Z_{g\alpha}^{r+1}$. In fact, for $t_\alpha^{r+1} \in C^\alpha$ we have

$$\nabla Ef_{0\alpha}^r(t_\alpha^{r+1}) = 0,$$

since $Ef_{0\alpha}^r \in L_\alpha^r$ and, consequently, $\nabla Ef_{0\alpha}^r \in L_\alpha^{r+1}$, and for $t_\alpha^{r+1} \in K_0^\alpha$,

$$\nabla Ef_{0\alpha}^r(t_\alpha^{r+1}) = \nabla f_{0\alpha}^r(t_\alpha^{r+1}) = 0,$$

since $f_{0\alpha}^r \in Z_{0\alpha}^r$.

4.3. We proceed now to prove Theorem 4.1.

To each $f_{0\alpha}^r \in Z_{0\alpha}^r$ corresponds a definite $g_\alpha^{r+1} \in Z_{g\alpha}^{r+1}$, namely $\nabla Ef_{0\alpha}^r$.

4.31. *The operator ∇E, applied to the group $Z^r_{0\alpha}$, generates an homomorphism ϕ_α of the group $B^r_{0\alpha}$ into the group $B^{r+1}_{g\alpha}$.*

In order to prove this it is sufficient to show that *if $f^r_{0\alpha} \varepsilon H^r_{0\alpha}$ then $\nabla E^r_{0\alpha} \varepsilon H^{r+1}_{g\alpha}$.*

Let $f^r_{0\alpha} \varepsilon H^r_{0\alpha}$; then

$$\overset{r}{f}_{0\alpha} = \nabla \overset{r-1}{f}_{0\alpha}.$$

Put

$$\overset{r-1}{f}_{\alpha} = E\overset{r-1}{f}_{0\alpha}. \tag{4.31}$$

For $t^r_{0\alpha} \varepsilon K^\alpha_0$ we have

$$\overset{r}{f}_{0\alpha}(\overset{r}{t}_{0\alpha}) = \Delta \overset{r-1}{f}_{\alpha}(\overset{r}{t}_{0\alpha}). \tag{4.32}$$

Put

$$\overset{r}{h}_\alpha(\overset{r}{t}_\alpha) = 0, \text{ if } \overset{r}{t}_\alpha \varepsilon \overset{\alpha}{K}_0, \tag{4.33}$$

$$\overset{r}{h}_\alpha(\overset{r}{t}_\alpha) = \nabla \overset{r-1}{f}_\alpha(\overset{r}{t}_\alpha), \text{ if } \overset{r}{t}_\alpha \varepsilon \overset{\alpha}{K} - \overset{\alpha}{K}_0. \tag{4.34}$$

From (4.32), (4.33), (4.34) follows

$$\overset{r}{h}_\alpha(\overset{r}{t}_\alpha) + E\overset{r}{f}_{0\alpha}(\overset{r}{t}_\alpha) = \nabla \overset{r-1}{f}_\alpha(\overset{r}{t}_\alpha)$$

for any $t^r_\alpha \varepsilon K^\alpha$, whence

$$\nabla \overset{r}{h}_\alpha + \nabla E\overset{r}{f}_{0\alpha} = 0, \quad \nabla E\overset{r}{f}_{0\alpha} = -\nabla \overset{r}{h}_\alpha,$$

i.e.,

$$\nabla E\overset{r}{f}_{0\alpha} \varepsilon H^{r+1}_{g\alpha}.$$

Let us now show that for any $u^r_\alpha \varepsilon B^r_{0\alpha}$

$$\overset{\alpha}{\pi}_\beta \phi_\alpha \overset{r}{u}_\alpha = \phi_\beta \overset{\alpha}{\pi}_\beta \overset{r}{u}_\alpha. \tag{4.35}$$

Let there be chosen some projection S^β_α of the complex K^β_α into the complex K^α. Choose arbitrarily $f^r_{0\alpha} \varepsilon u^r_\alpha$. In order to prove the formula (4.35) it is sufficient to show that under the homomorphism $\sigma^{0\alpha}_{0\beta}$ of the group $Z^r_{0\alpha}$ into the group $Z^r_{0\beta}$ generated by the projection S^β_α we have

$$\overset{\alpha}{\sigma}_\beta \nabla E\overset{r}{f}_{0\alpha} \sim \nabla E\overset{\alpha}{\sigma}_\beta \overset{r}{f}_{0\alpha} \text{ in } \overset{\beta}{K} - \overset{\beta}{K}_0 - \overset{\beta}{C}. \tag{4.351}$$

Having in view that

$$\overset{\alpha}{\sigma}_\beta \nabla = \nabla \overset{\alpha}{\sigma}_\beta,$$

we may write (4.351) in the form

$$\nabla \overset{\alpha}{\sigma}_\beta E\overset{r}{f}_{0\alpha} - \nabla E\overset{\alpha}{\sigma}_\beta \overset{r}{f}_{0\alpha} \sim 0 \qquad \text{in } \overset{\beta}{K} - \overset{\beta}{K}_0 - \overset{\beta}{C}. \tag{4.352}$$

But (4.352) obviously follows from

$$(4.353) \qquad \overset{\alpha}{\sigma_\beta} E \overset{r}{f_{0\alpha}} - E \overset{\alpha}{\sigma_\beta} \overset{r}{f_{0\alpha}} \varepsilon \overset{r}{L_{\varrho\beta}}.$$

We have thus only to verify this last formula. But

$$\overset{\alpha}{\sigma_\beta} E \overset{r}{f_{0\alpha}}(\overset{r}{t_\beta}) = E \overset{r}{f_{0\alpha}}(\overset{\beta}{S_\alpha} \overset{r}{t_\beta}),$$

that is,

$$\begin{aligned} \overset{\alpha}{\sigma_\beta} E \overset{r}{f_{0\alpha}}(\overset{r}{t_\beta}) &= 0, \qquad \text{if } \overset{\beta}{S_\alpha} \overset{r}{t_\beta} \varepsilon \overset{\alpha}{C}, \\ &= \overset{r}{f_{0\alpha}}(\overset{\beta}{S_\alpha} \overset{r}{t_\beta}), \text{ if } \overset{\beta}{S_\alpha} \overset{r}{t_\beta} \varepsilon \overset{\alpha}{K_0}, \\ &= 0, \qquad \text{if } \overset{\beta}{S_\alpha} \overset{r}{t_\beta} \varepsilon \overset{\alpha}{K} - \overset{\alpha}{K_0}. \end{aligned}$$

On the other hand,

$$\begin{aligned} E \overset{\alpha}{\sigma_\beta} \overset{r}{f_{0\alpha}}(\overset{r}{t_\beta}) &= 0, \qquad \overset{r}{t_\beta} \varepsilon \overset{\beta}{C}, \\ E \overset{\alpha}{\sigma_\beta} \overset{r}{f_{0\alpha}}(\overset{r}{t_\beta}) &= \overset{\alpha}{\sigma_\beta} \overset{r}{f_{0\alpha}}(\overset{r}{t_\beta}) = \overset{r}{f_{0\alpha}}(\overset{\beta}{S_\alpha} \overset{r}{t_\beta}), \text{ if } \overset{r}{t_\beta} \varepsilon \overset{\beta}{K_0}, \\ E \overset{\alpha}{\sigma_\beta} \overset{r}{f_{0\alpha}}(\overset{r}{t_\beta}) &= 0, \text{ if } \overset{r}{t_\beta} \varepsilon \overset{\beta}{K} - \overset{\beta}{K_0}. \end{aligned}$$

If $t^r_\beta \varepsilon K^\beta_0$ or $t^r_\beta \varepsilon C^\beta$, then, respectively,

$$\overset{\beta}{S_\alpha} \overset{r}{t_\beta} \varepsilon \overset{\alpha}{K_0}, \qquad \overset{\beta}{S_\alpha} \overset{r}{t_\beta} \varepsilon \overset{\alpha}{C}.$$

Therefore for

$$\overset{r}{t_\beta} \varepsilon \overset{\beta}{K_0}, \qquad \overset{r}{t_\beta} \varepsilon \overset{\beta}{C},$$

we have

$$\overset{\alpha}{\sigma_\beta} E \overset{r}{f_{0\alpha}}(\overset{r}{t_\beta}) = E \overset{\alpha}{\sigma_\beta} \overset{r}{f_{0\alpha}}(\overset{r}{t_\beta}),$$

whence follows (4.353) and so (4.352), (4.351) and (4.35).

From (4.35) and (3.41) follows

4.32. *The operator ∇E determines through the homomorphisms ϕ_α the homomorphism ϕ of the group B^r_0 into the group B^{r+1}_ϱ.*

If u^r is an element of the group B^r_0, then take any element u^r_α of the bundle u^r and any cycle f^r_α contained in the homologic class of u^r. The element u^{r+1}_α of the group $B^{r+1}_{\varrho\alpha}$ containing the cycle $\nabla E f^r_\alpha$ is contained in the bundle v^{r+1}, which is by definition the element $\phi(u^r)$ of the group B^{r+1}_ϱ.

4.4. The homomorphism ϕ is a mapping of the group B^r_0 *on* the group B^{r+1}_ϱ. In order to prove this it is, on ground of (3.411), sufficient to show that

4.41. *Whatever be α and $v^{r+1}_\alpha \varepsilon B^{r+1}_{\varrho\alpha}$, there is always a $K^\beta > K^\alpha$ and a $u^r_\beta \varepsilon B^r_{0\beta}$ such that*

$$\phi_\beta \overset{r}{u_\beta} = \overset{\alpha}{\pi_\beta} \overset{r+1}{v_\alpha}.$$

Let v_α^{r+1} be given. Take a $g_\alpha^{r+1} \varepsilon v_\alpha^{r+1}$. Since K is simply connected with respect to the dimensionality $r+1$, there is always a $K^\beta > K^\alpha$ such that

$$\overset{\alpha}{\sigma}_\beta \overset{r+1}{g_\alpha} \sim 0 \qquad \text{in } K^\beta - C^\beta.$$

Thus there exists such an f_β^r that

$$\nabla \overset{r}{f_\beta} = \overset{\alpha}{\sigma}_\beta \overset{r+1}{g_\alpha}. \tag{4.41}$$

Putting

$$\overset{\alpha}{\sigma}_\beta \overset{r+1}{g_\alpha} = \overset{r+1}{g_\beta}, \tag{4.42}$$

we have

$$\overset{r+1}{g_\beta} \varepsilon \overset{r+1}{v_\beta} \varepsilon \overset{\alpha}{\pi}_\beta \overset{r+1}{v_\alpha}. \tag{4.43}$$

If $t_\beta^{r+1} \varepsilon K_0^\beta$, then

$$\nabla A \overset{r}{f_\beta}(\overset{r+1}{t_\beta}) = \nabla \overset{r}{f_\beta}(\overset{r+1}{t_\beta}) = 0,$$

i.e.,

$$A \overset{r}{f_\beta} \varepsilon \overset{r}{Z}_{0\beta}.$$

Observing that always $EAf_\beta^r = \bar{f}_\beta^r$, put

$$\overset{r}{g_\beta} = \overset{r}{f_\beta} - EA\overset{r}{f} = \overset{r}{f_\beta} - \overset{r}{\bar{f}_\beta} \varepsilon \overset{r}{L}_{0\beta}.$$

Evidently

$$\nabla \overset{r}{g_\beta} = \nabla \overset{r}{f_\beta} - \nabla EA\overset{r}{f_\beta}.$$

Consequently, in view of (4.41) and (4.42),

$$\overset{r+1}{g_\beta} \sim \Delta EA\overset{r}{f_\beta} \qquad \text{in } K^\beta - K_0^\beta - C^\beta,$$

i.e. (on ground of (4.43)),

$$\nabla EA\overset{r}{f_\beta} \varepsilon \overset{\alpha}{\pi}_\beta \overset{r+1}{v_\alpha}.$$

If $Af_\beta^r \varepsilon u_\beta^r \varepsilon B_\beta^r$, then

$$\phi_\beta \overset{r}{u_\beta} = \overset{\alpha}{\pi}_\beta \overset{r+1}{v_\alpha},$$

and 4.41 is proved.

4.5. *The homomorphism ϕ is an isomorphism.*

LEMMA 4.51. *If $f_\alpha^r \varepsilon Z_\nabla^r(K^\alpha)$, then there exists such a $K^\beta > K^\alpha$ that*

$$A\overset{\alpha}{\sigma}_\beta \overset{r}{f_\alpha} \sim 0 \qquad \text{in } K_0^\beta - C^\beta.$$

In fact, since K is simply-connected with respect to the dimensionality r, in some $K^\beta > K^\alpha$ there exists such an $f_\beta^{r-1} \subset K^\beta - C^\beta$ that

$$\sigma_\beta^\alpha f_\alpha^r = \nabla f_\beta^{r-1}.$$

Hence

$$A\sigma_\beta^\alpha f_\alpha^r = A\nabla f_\beta^{r-1} = \nabla A f_\beta^{r-1},$$

and Lemma 4.51 is proved.

For the proof of the assertion 4.5 it is sufficient, in virtue of (3.412), to show that if

$$\phi u^r = 0 \tag{4.52}$$

and

$$f_{0\alpha}^r \,\varepsilon\, u_\alpha^r \,\varepsilon\, u^r \tag{4.53}$$

then there exists such a $K^\beta > K^\alpha$ that

$$\sigma_\beta^\alpha f_{0\alpha}^r \sim 0 \qquad \text{in } K_0^\beta - C^\beta. \tag{4.54}$$

From (4.52) it follows that (4.53) may be from the outset chosen in such a way that

$$\nabla E f_{0\alpha}^r = \nabla g_\alpha^r, \qquad g_\alpha^r \,\varepsilon\, L_{g\alpha}^r.$$

Evidently $Ef_{0\alpha}^r - g_\alpha^r$ is a cycle and hence, in virtue of Lemma 4.51, we may choose $K^\beta > K^\alpha$ such that

$$A(\sigma_\beta^\alpha E f_{0\alpha}^r - \sigma_\beta^\alpha g_\alpha^r) \sim 0 \qquad \text{in } K_0^\beta - C^\beta.$$

Since

$$A\sigma_\beta^\alpha g_\alpha^r = 0,$$

$$A\sigma_\beta^\alpha E f_{0\alpha}^r \sim 0 \qquad \text{in } K_0^\beta - C^\beta,$$

and since

$$A\sigma_\beta^\alpha E f_{0\alpha}^r = \sigma_\beta^\alpha A E f_{0\alpha}^r = \sigma_\beta^\alpha f_{0\alpha}^r,$$

$$\sigma_\beta^\alpha f_{0\alpha}^r \sim 0 \qquad \text{in } K_0^\beta - C^\beta,$$

so that the assertion 4.5 and with it the whole Theorem 4.41 is completely proved.

4.6. Consider separately the case $r=0$. Suppose that all K^α are connected complexes, so that the groups $B_\nabla^{00}(K^\alpha)$ consist only of the corresponding zero elements.

Suppose, further, that all $C^\alpha = 0$ and that our system of complexes $\mathfrak{K}$ is simply-connected with respect to the dimensionality 1.

Instead of $B^{00}_{\nabla}(K^{\alpha}_{0})$ we shall write $B^{00}_{0\alpha}$ and instead of lim $[B^{00}_{0\alpha}; \pi^{\alpha}_{\beta}]$ we shall write simply B^{00}.

The operator ∇E again correlates to every element $f^{r}_{0\alpha} \varepsilon Z^{0}_{0\alpha}$ the element

$$\overset{1}{g}_{\alpha} = \nabla E \overset{r}{f}_{0\alpha} \varepsilon \overset{1}{Z}_{g}.$$

Moreover, *if* $f^{0}_{0\alpha} \varepsilon Z^{00}_{0\alpha}$, *then* $\nabla E f^{0}_{0\alpha} \varepsilon H^{1}_{g\alpha}$. In fact, putting the function f^{0}_{α} on the whole K^{α} equal to the constant value of the function $f^{0}_{0\alpha}$ on K^{α}_{0}, we obtain on ground of 1.81 a cycle f^{0}_{α}. Here

$$\overset{0}{f}_{\alpha} = E\overset{0}{f}_{0\alpha} + \overset{0}{h}_{\alpha},$$

where

$$\overset{0}{h}_{\alpha} = \overset{0}{f}_{\alpha} \text{ on } K^{\alpha} - K^{\alpha}_{0}, \qquad \overset{0}{h}_{\alpha} = 0 \text{ on } K^{\alpha}_{0}.$$

Since f^{0}_{α} is a cycle,

$$0 = \nabla \overset{0}{f}_{\alpha} = \nabla E \overset{0}{f}_{0\alpha} + \nabla \overset{0}{h}_{\alpha},$$

i.e.,

$$\nabla E \overset{0}{f}_{0\alpha} \varepsilon \overset{0}{H}_{g\alpha}.$$

Thus *the operator* ∇E *applied to the group* $Z^{0}_{0\alpha}$ *generates a homomorphism* ϕ_{α} *of the group* $B^{00}_{0\alpha}$ *into the group* $B^{1}_{g\alpha}$.

Similarly, as in 4.3 we prove the formula (4.351), where now $r=0$ and the arguments are only simplified by the fact that $C^{\alpha}=0$; from (4.351) follows (4.35), where $r=0$ and u^{r}_{α} denotes an arbitrary element of the group $B^{00}_{0\alpha}$. Thus *the operator* ∇E *determines through the homomorphisms* ϕ_{α} *a homomorphism* ϕ *of the group* B^{00}_{0} *into the group* B^{1}_{g}.

The reasonings of 4.4 remain in force and prove that the homomorphism ϕ is a mapping of the group B^{00}_{0} on the group B^{1}_{g}.

Let us finally prove that ϕ is an isomorphism. From the connectivity of K^{α} and from 1.812 follows in the first place:

LEMMA. *If* f^{0}_{α} *is a cycle, then*

$$Af_{\alpha} \varepsilon \overset{00}{Z}_{\alpha}.$$

In order to prove that ϕ is an isomorphism, it is sufficient to prove that if $u^{0} \varepsilon B^{00}_{0}$ and, further,

$$\phi(u_{0}) = 0, \tag{4.61}$$

$$f_{0\alpha} \varepsilon \overset{0}{u}_{\alpha} \varepsilon \overset{0}{u}, \tag{4.62}$$

then

$$\overset{0}{f}_{0\alpha} \varepsilon \overset{00}{Z}_{0\alpha}. \tag{4.63}$$

From (4.61) it follows that (4.62) may be from the outset so chosen that

$$\nabla E\overset{0}{f}_{0\alpha} = \nabla \overset{0}{g}_{\alpha}; \qquad \overset{0}{g}_{\alpha} \varepsilon \overset{0}{L}(K^{\alpha} - K^{\alpha}_{0}).$$

Evidently $Ef^0_{0\alpha} - g^0_\alpha$ is a cycle and hence, according to the above lemma,

$$A(E\overset{0}{f}_{0\alpha} - \overset{0}{g}_{\alpha}) \varepsilon \overset{00}{Z}_{0\alpha},$$

i.e.,

$$AE\overset{0}{f}_{0\alpha} \varepsilon \overset{00}{Z}_{0\alpha},$$

i.e.,

$$\overset{0}{f}_{0\alpha} \varepsilon \overset{00}{Z}_{0\alpha}.$$

Thus we have proved

THEOREM 4.6. *If there is given an unbounded partially ordered set of complexes K^α connected and simply-connected with respect to the dimensionality* 1 *and their subcomplexes K^α satisfying the conditions of* 4.1 *for $C^\alpha = 0$, then the groups*

$$\lim\,[\overset{00}{B}_\nabla(K^{\alpha}_{0}); \overset{\alpha}{\pi}_{\beta}], \qquad \lim\,[\overset{1}{B}_\nabla(K^{\alpha} - K^{\alpha}_{0}); \overset{\alpha}{\pi}_{\beta}]$$

are isomorphic.

4.7. From Theorem 4.1 follows

4.71. *If C^α and K^α_0 have no common elements and each of the groups*

$$\lim\,[\overset{r}{B}_\nabla(K^{\alpha} - C^{\alpha}); \overset{\alpha}{\pi}_{\beta}], \qquad \lim\,[\overset{r+1}{B}_\nabla(K^{\alpha} - C^{\alpha}); \overset{\alpha}{\pi}_{\beta}],$$
$$\lim\,[\overset{r}{B}_\nabla(K^{\alpha}); \overset{\alpha}{\pi}_{\beta}], \qquad \lim\,[\overset{r+1}{B}_\nabla(K^{\alpha}); \overset{\alpha}{\pi}_{\beta}]$$

contains only the zero element, then the groups

$$\overset{r+1}{B}_{s} = \lim\,[\overset{r+1}{B}(K^{\alpha} - K^{\alpha}_{0} - C^{\alpha}); \overset{\beta}{\pi}_{\alpha}]$$

and

$$\overset{r+1}{\mathfrak{B}}_{s} = \lim\,[\overset{r+1}{B}_\nabla(K^{\alpha} - K^{\alpha}_{0}); \overset{\alpha}{\pi}_{\beta}]$$

are isomorphic.

In fact, in our case $K^\alpha_0 - C^\alpha = K^\alpha_0$ and both groups B^{r+1}_s and $\mathfrak{B}^{r+1}_s$ are isomorphic to the group

$$\lim\,[\overset{r}{B}_\nabla(K^{\alpha}_{0}); \overset{\alpha}{\pi}_{\beta}].$$

4.8. *Remark.* In what follows we often will have to do with unbounded partially ordered sets $\mathfrak{K}$ of complexes K^α with given subcomplexes $C^\alpha \subset K^\alpha$, and of simplicial mappings S^β_α of K^β into K^α ("projections"), defined for every $K^\beta > K^\alpha$. These projections will satisfy the conditions 1° and 2° of §4.1, with

respect to C^α and the condition 3°, so that there is a direct spectrum

$$[B^r_\nabla(K^\alpha - C^\alpha);\ \pi^\alpha_\beta],$$

π^α_β being the homomorphism of $B^r_\nabla(K^\alpha - C^\alpha)$ into $B^r_\nabla(K^\beta - C^\beta)$, generated by S^β_α. Under these circumstances the limit-group

$$B^r_\nabla = \lim\ [B^r_\nabla(K^\alpha - C^\alpha);\ \pi^\alpha_\beta]$$

can be defined in a particularly simple way. In fact, we shall consider the set Z^r of elements of all groups $Z^r_\nabla(K^\alpha - C^\alpha)$, i.e.,

$$Z^r = \cup_\alpha Z^r(K^\alpha - C^\alpha)$$

and call $z^r_\alpha \ \varepsilon\ Z^r(K^\alpha - C^\alpha)$ and $z^r_\beta \ \varepsilon\ Z^r(K^\beta - C^\beta)$ equivalent if a K^γ,

$$K^\gamma > K^\alpha, \qquad K^\gamma > K^\beta$$

and projections S^γ_α, S^γ_β of respectively K^γ and K^β into K^α can be found in such a way that

$$\overset{\alpha}{\sigma}_\gamma \overset{r}{z}_\alpha \sim \overset{\beta}{\sigma}_\gamma \overset{r}{z}_\beta \qquad \text{in } K^\gamma - C^\gamma,$$

σ^α_γ, σ^β_γ being the homomorphisms of $Z^r(K^\alpha - C^\alpha)$ and $Z^r(K^\beta - C^\beta)$ in $Z^r(K^\gamma - C^\gamma)$ generated by S^γ_α and S^γ_β respectively. Thus Z^r is divided into classes or *bundles* of equivalent cycles and these bundles form a group, which, by definition, is the group B^r_∇; the addition in B^r_∇ is defined in the following way: ζ^r_1 and ζ^r_2 being two bundles, we take an $z^r_\alpha \ \varepsilon\ \zeta^r_1$ and a $z'^r_\alpha \ \varepsilon\ \zeta^r_2$ and call $\zeta^r_1 + \zeta^r_2$ the bundle containing $z^r_\alpha + z'^r_\alpha$. It is an easy task to show that this definition of

$$\lim\ [B^r_\nabla(K^\alpha - C^\alpha);\ \pi^\alpha_\beta]$$

agees with the definition given in §3.2.

5. Coverings

5.1. In the present paragraph R denotes one and the same infinite set.

An *indexed subset* e_i is by definition a pair consisting of a certain subset $|e_i|$ of the set R and a natural number i. Two indexed subsets e_i and e_j are considered to be equal if the sets $|e_i|$ and $|e_j|$ are identical and the indices i and j are equal.

By a *covering* of a set we understand such a finite system of indexed subsets

$$\Omega = \{e_1, \cdots, e_s\} \tag{5.1}$$

that

$$|e_1| \cup \cdots \cup |e_s| = R.$$

A covering Ω is called *simple* if the identity $|e_i| = |e_j|$ is realized only in the case of equality $i = j$, i.e., only in the case of equality $e_i = e_j$. Since in the

case of a simple covering the indexed sets e_i correspond to the sets $|e_i|$ in one-to-one manner, we may not distinguish between the first and the latter and consider the sets $|e_i|$ themselves to be the elements of the simple covering.

The nerve of the covering $\Omega=\{e_1, \cdots, e_s\}$ is the complex K with the vertices $e_1, \cdots, e_s$; the vertices

$$e_{i_0}, \cdots, e_{i_r}$$

form a simplex of the complex K when and only when

$$|e_{i_0}| \cap \cdots \cap |e_{i_r}| \neq 0.$$

If a covering is denoted by the letter Ω with some indices, then we shall denote its nerve by the letter K with the same indices. For instance, the nerve of the covering Ω^α shall be denoted by K^α, the nerve of the covering $\Omega^{\alpha\lambda}$ by $K^{\alpha\lambda}$, etc.

5.2. DEFINITION 5.21. *A covering*

$$\Omega^\beta = \{e_j^\beta\}, \qquad j = 1, \cdots, s_\beta,$$

is called a subdivision of the covering

$$\Omega^\alpha = \{e_i^\alpha\}, \qquad i = 1, \cdots, s_\alpha,$$

if each of the $|e_0^\beta|$ *is contained in at least one of the* $|e_i^\alpha|$.

DEFINITION 5.22. *A covering* Ω^β *follows after the covering* Ω^α,

$$\Omega^\beta > \Omega^\alpha,$$

if Ω^β *is a subdivision of the covering* Ω^α, *but* Ω^α *is not a subdivision of the covering* Ω^β.

This definition of "follows" turns the set of all coverings Ω of a set R into a partially ordered set $\mathfrak{W}$. Every part $\mathfrak{B}$ of the partially ordered set $\mathfrak{W}$ is called a *system of coverings of the set R*.

From Definition 5.22 follows

5.221. *If* $\Omega^\beta > \Omega^\alpha$ *and* Ω^γ *is a subdivision of* Ω^β, *then* $\Omega^\gamma > \Omega^\alpha$; *if* Ω^β *is a subdivision of* Ω^α *and* $\Omega^\gamma > \Omega^\beta$, *then* $\Omega^\gamma > \Omega^\alpha$.

For the nerves of the coverings we put $K^\beta > K^\alpha$, if $\Omega^\beta > \Omega^\alpha$.

Let a covering $\Omega^\alpha = \{e_i^\alpha\}$ and its subdivision $\Omega^\beta = \{e_j^\beta\}$ be given. To each e_j^β we correlate some definite e_i^α under the only condition that

$$|e_j^\beta| \subset |e_i^\alpha|.$$

Such a mapping of the covering Ω^β into the covering Ω^α and also the corresponding simplicial mapping of the complex K^β into the complex K^α shall be

called a *protection* and denoted by S_α^β or, if no misunderstandings are to be feared, simply by S.

5.3. Suppose that a certain system $\mathfrak{C}$ of subsets of the set R called *special* subsets is singled out. We suppose that the system of special subsets satisfies the following condition:

5.31. *If E is a special subset and $E' \supset E$, then E' is also a special subset.*

The element e_i^α of the covering Ω^α is called *special*, if $|e_i^\alpha|$ is a special subset.

Special elements of the covering Ω^α define a certain subcomplex C^α of the complex K^α called the *special subcomplex* of the complex K^α.

From the condition 5.31 follows

5.32. *For any protection S_α^β of the covering Ω^β into the covering Ω^α we have* $S_\alpha^\beta C^\beta \subset C^\alpha$.

Let S and $\tilde{S}$ be two projections of the covering Ω^β into the covering Ω^α. If

$$t_\beta = (e_{i_0}, \cdots, e_{i_r})$$

is any simplex from K^β, then the vertices

$$Se_{i_0}, \tilde{S}e_{i_0}; \cdots ; Se_{i_r}, \tilde{S}e_{i_r}$$

define a simplex t_α in K^α having among its faces St_β as well as $\tilde{S}t_\beta$; if, moreover, $t_\beta \in C^\beta$, then $t_\alpha \in C^\alpha$. Hence all projections of the covering Ω^β into the covering Ω^α define one and the same homomorphism $\tilde{\omega}_\alpha^\beta$ of the group $B_\Delta^r(K^\beta \bmod C^\beta)$ into the group $B_\Delta^r(K^\alpha \bmod C^\alpha)$ and one and the same homomorphism π_β^α of the group $B_\nabla^r(K^\alpha - C^\alpha)$ into the group $B_\nabla^r(K^\beta - C^\beta)$.

If the system of coverings $\mathfrak{B}$ is unbounded, then we obtain the inverse and the direct spectra

$$[B_\Delta^r(K^\alpha \bmod C^\alpha); \tilde{\omega}_\alpha^\beta], \qquad [B_\nabla^r(K^\alpha - C^\alpha); \pi_\beta^\alpha],$$

the limit groups of which we denote respectively by $B_\Delta^r(\mathfrak{B}, \mathfrak{C})$ and $B_\nabla^r(\mathfrak{B}, \mathfrak{C})$. From the investigations of Steenrod(4) it follows that the groups $B_\Delta^r(\mathfrak{B}, \mathfrak{C})$ and $B_\nabla^r(\mathfrak{B}, \mathfrak{C})$ are dual to each other.

5.4. Consider some covering, which we shall denote by

$$\Omega^{\alpha\lambda} = \{e_{ik}\};$$

by $e_{i1}, e_{i2}, \cdots, e_{ih_i}$ are denoted all elements of the covering $\Omega^{\alpha\lambda}$ for which $|e_{i1}| = |e_{i2}| = \cdots = |e_{ih_i}|$, namely $|e_{ik}| = e_i$. The sets e_i form a simple covering Ω^α denoted also by $|\Omega^{\alpha\lambda}|$.

To every element e_{ik} of the covering $\Omega^{\alpha\lambda}$ there corresponds an element e_i of the covering Ω^α and this correspondence establishes a simplicial mapping $D_\alpha^{\alpha\lambda}$ of the complex $K^{\alpha\lambda}$ into the complex K^α, under which the special complex

$C^{\alpha\lambda} \subset K^{\alpha\lambda}$ is transformed into the special complex $C^{\alpha} \subset K^{\alpha}$. The simplicial mapping $D^{\alpha\lambda}_{\alpha}$ generates a homomorphism $\rho^{\alpha\lambda}_{\alpha}$ of the group $B^r_{\Delta}(K^{\alpha\lambda} \bmod C^{\alpha\lambda})$ into the group $B^r_{\Delta}(K^{\alpha} \bmod C^{\alpha})$ and a homomorphism $\sigma^{\alpha}_{\alpha\lambda}$ of the group $B^r_{\nabla}(K^{\alpha} - C^{\alpha})$ into the group $B^r_{\nabla}(K^{\alpha\lambda} - C^{\alpha\lambda})$.

5.41. *The homomorphisms $\rho^{\alpha\lambda}_{\alpha}$ and $\sigma^{\alpha}_{\alpha\lambda}$ are isomorphic mappings of, correspondingly,* $B^r_{\Delta}(K^{\gamma\lambda} \bmod C^{\alpha\lambda})$ *on* $B^r_{\Delta}(K^{\alpha} \bmod C^{\alpha})$ *and of* $B^r_{\nabla}(K^{\alpha} - C^{\alpha})$ *on* $B^r_{\nabla}(K^{\alpha\lambda} - C^{\alpha\lambda})$.

For the proof observe in the first place that the subcomplex $K^{\alpha\lambda 1}$ of the complex $K^{\alpha\lambda}$ consisting of all simplexes of $K^{\alpha\lambda}$ whose vertices have the form e_{i1} is isomorphic to the complex K^{α}. In virtue of this isomorphism between $K^{\alpha\lambda 1}$ and K^{α} to the simplicial mapping $D^{\alpha\lambda}_{\alpha}$ of the complex $K^{\alpha\lambda}$ on the complex K^{α} there corresponds a simplicial mapping $D^{\alpha\lambda}_{\alpha\lambda 1}$ of the complex $K^{\alpha\lambda}$ on $K^{\alpha\lambda 1}$ correlating to the vertex e_{ik} the vertex e_{i1} and, consequently, leaving all vertices and all simplexes of the complex $K^{\alpha\lambda 1} \subset K^{\alpha\lambda}$ fixed.

An arbitrary simplex

$$t_{\alpha\lambda} = (e_{i_0 k_0}, \cdots, e_{i_r k_r})$$

of the complex $K^{\alpha\lambda}$ and its image under the mapping $D^{\alpha\lambda}_{\alpha\lambda 1}$,

$$D^{\alpha\lambda}_{\alpha\lambda 1}(t_{\alpha\lambda}) = (e_{i_0 1}, \cdots, e_{i_r 1})$$

are faces of a simplex

$$T_{\alpha\lambda} = (e_{i_0 1}, \cdots, e_{i_0 h_0}, \cdots, e_{i_r 1}, \cdots, e_{i_r h_r})$$

(we write h_0 instead of $h_{i_0}, \cdots, h_r$ instead of h_{i_r}) belonging to the complex $K^{\alpha\lambda}$ and, moreover, if $t_{\alpha\lambda} \in C^{\alpha\lambda}$, then $T_{\alpha\lambda} \in C^{\alpha\lambda}$. Hence it follows that for

$$\overset{r}{f}_{\alpha\lambda} \in \overset{r}{Z}_{\Delta}(K^{\alpha\lambda} \bmod C^{\alpha\lambda})$$

we have

$$D^{\alpha\lambda}_{\alpha\lambda 1}\overset{r}{f}_{\alpha\lambda} \sim \overset{r}{f}_{\alpha\lambda} \text{ modulo } C^{\alpha\lambda} \qquad \text{in } K^{\alpha\lambda},$$

i.e., for every homologic class

$$\overset{r}{\zeta}_{\alpha\lambda} \in B^r_{\Delta}(K^{\alpha\lambda} \bmod C^{\alpha\lambda})$$

we have

$$\rho^{\alpha\lambda}_{\alpha\lambda 1}\overset{r}{\zeta}_{\alpha\lambda} \subset \overset{r}{\zeta}_{\alpha\lambda}.$$

On the other hand every homologic class

$$\overset{r}{\zeta}_{\alpha\lambda 1} \in B^r_{\Delta}(K^{\alpha\lambda 1} \bmod C^{\alpha\lambda 1})$$

is contained in a uniquely determined homologic class

$$\overset{r}{\zeta}_{\alpha\lambda} \in B^r_{\Delta}(K^{\alpha\lambda} \bmod C^{\alpha\lambda}).$$

It remains to show that the homologic class $\zeta^r_{\alpha\lambda}$ contains only one homologic class $\zeta^r_{\alpha\lambda 1}$, i.e., that from

$$\overset{r}{f}_{\alpha\lambda 1} \varepsilon Z^r_\Delta(K^{\alpha\lambda 1} \bmod C^{\alpha\lambda 1}), \qquad \overset{r}{f}_{\alpha\lambda 1} \varepsilon H^r_\Delta(K^{\alpha\lambda} \bmod C^{\alpha\lambda})$$

follows

$$\overset{r}{f}_{\alpha\lambda 1} \varepsilon H^r_\Delta(K^{\alpha\lambda 1} \bmod C^{\alpha\lambda 1}). \tag{5.41}$$

Since under our assumptions

$$\overset{r}{f}_{\alpha\lambda 1} = \Delta \overset{r+1}{f}_{\alpha\lambda},$$

we have

$$\overset{r}{f}_{\alpha\lambda 1} = \overset{\alpha\lambda}{D}_{\alpha\lambda 1}\overset{r}{f}_{\alpha\lambda 1} = \overset{\alpha\lambda}{D}_{\alpha\lambda 1}\Delta \overset{r+1}{f}_{\alpha\lambda} = \Delta \overset{\alpha\lambda}{D}_{\alpha\lambda 1}\overset{r+1}{f}_{\alpha\lambda},$$

so that the inclusion (5.41) is proved.

Thus $\rho^{\alpha\lambda}_\alpha$ is an isomorphic mapping of the group $B^r_\Delta(K^{\alpha\lambda} \bmod C^{\alpha\lambda})$ on $B^r_\Delta(K^\alpha \bmod C^\alpha)$, consequently $\sigma^\alpha_{\alpha\lambda}$ is an isomorphic mapping of the group $B^r_\nabla(K^\alpha - C^\alpha)$ on $B^r_\nabla(K^{\alpha\lambda} - C^{\alpha\lambda})$.

5.5. Let there be given coverings $\Omega^{\alpha\lambda}$, Ω^α, $\Omega^{\beta\mu}$, Ω^β, connected by the relations

$$\Omega^\alpha = |\Omega^{\alpha\lambda}|, \qquad \Omega^\beta = |\Omega^{\beta\mu}|; \qquad \Omega^{\beta\mu} > \Omega^{\alpha\lambda}.$$

Then also $\Omega^\beta > \Omega^\alpha$ and, moreover, if $S^{\beta\mu}_{\alpha\lambda}$ is a projection of $\Omega^{\beta\mu}$ into $\Omega^{\alpha\lambda}$, then we have a completely determined projection S^β_α defined by the formula

$$S^\beta_\alpha \overset{\beta}{e}_j = S^\beta_\alpha D^{\beta\mu}_\beta \overset{\beta\mu}{e}_{j1} = D^{\alpha\lambda}_\alpha S^{\beta\mu}_{\alpha\lambda} \overset{\beta\mu}{e}_{j1}. \tag{5.51}$$

Denoting by $\rho^{\alpha\lambda}_\alpha$, $\rho^{\beta\mu}_\beta$, $\sigma^\alpha_{\alpha\lambda}$, $\sigma^\beta_{\beta\mu}$ the isomorphisms (defined in 5.4) generated by the mappings $D^{\alpha\lambda}_\alpha$ and $D^{\beta\mu}_\beta$, we see that $\bar{\omega}^\beta_\alpha \rho^{\beta\mu}_\beta$ and $\rho^{\alpha\lambda}_\alpha \bar{\omega}^{\beta\mu}_{\alpha\lambda}$ is one and the same homomorphism of the group $B^r_\Delta(K^{\beta\mu} \bmod C^{\beta\mu})$ into the group $B^r_\Delta(K^\alpha \bmod C^\alpha)$ and, consequently, $\sigma^\beta_{\beta\mu}\pi^\alpha_\beta$ and $\pi^{\alpha\lambda}_{\beta\mu}\sigma^\beta_{\beta\mu}$ express one and the same homomorphism of the group $B^r_\nabla(K^\alpha - C^\alpha)$ into the group $B^r_\nabla(K^{\beta\mu} - C^{\beta\mu})$. Hence

$$\overset{\alpha\lambda}{\pi}_{\beta\mu} = \overset{\beta}{\sigma}_{\beta\mu} \overset{\alpha}{\pi}_\beta (\overset{\alpha}{\sigma}_{\alpha\lambda})^{-1},$$

or, denoting by $\phi^{\alpha\lambda}_\alpha$, $\phi^{\beta\mu}_\mu$ the isomorphic mappings respectively inverse to the isomorphisms $\sigma^\alpha_{\alpha\lambda}$ and $\sigma^\beta_{\beta\mu}$,

$$\overset{\alpha\lambda}{\pi}_{\beta\mu} = (\overset{\beta\mu}{\phi}_\beta)^{-1} \overset{\alpha}{\pi}_\beta \overset{\alpha\lambda}{\phi}_\alpha. \tag{5.52}$$

5.6. Suppose now that we have an unbounded system $\mathfrak{B}$ of coverings $\Omega^{\alpha\lambda}$ of the set R. Let the system

$$|\mathfrak{B}| = \{\Omega^\alpha\},$$

where $\Omega^\alpha = |\Omega^{\alpha\lambda}|$ be also unbounded. Under these conditions we have

5.61. *The groups*

$$B_\Delta^r(\mathfrak{B}, \mathfrak{C}) = \lim\ [B_\nabla^r(K^{\alpha\lambda} - C^{\alpha\lambda}); \pi_{\beta\mu}^{\alpha\lambda}],$$
$$B_\nabla^r(|\mathfrak{B}|, \mathfrak{C}) = \lim\ [B_\nabla^r(K^\alpha - C^\alpha); \pi_\beta^\alpha]$$

are isomorphic.

In fact, in virtue of (5.52) the conditions of 3.61 are satisfied.

Let us establish an important particular case of 5.61.

Let A be any subset of the set R. Suppose that, the condition 5.31 being satisfied, some subsets of the set A are singled out as special. Denote the system of special subsets of the set A by $\mathfrak{C}_0$.

Let there be given an unbounded system $\mathfrak{B}$ of coverings Ω^α of the set R. We construct for each $\Omega^\alpha = \{e_i^\alpha\}$ a covering $A\,\Omega^\alpha$ of the set A in the following manner. The elements of the covering $A\,\Omega^\alpha$ are the indexed sets

$$(A \cap |e_i^\alpha|)_i,$$

which we shall simply denoted by Ae_i^α.

If $\Omega^\beta > \Omega^\alpha$, put $A\,\Omega^\beta > A\,\Omega^\alpha$. The obtained system of coverings $A\,\Omega^\alpha$ we shall denote by $A\mathfrak{B}$.

The system $|A\mathfrak{B}|$ consists of all simple coverings $|A\,\Omega^\alpha|$, where $|A\,\Omega^\alpha|$ may be defined as the simple covering consisting of all non-void sets representable in the form $A \cap |e_i^\alpha|$, where $e_i^\alpha \in \Omega^\alpha$.

From what has been proved above follows

5.611. *If for an unbounded system of coverings V of the set R and for a subset $A \subset R$ the system $|A\mathfrak{B}|$ is also unbounded, then the groups $B_\nabla^r(A\,V, \mathfrak{C}_0)$ and $B_\nabla^r(|A\mathfrak{B}|, \mathfrak{C}_0)$ are isomorphic.*

6. Coverings of topological spaces

6.1. A covering of a topological space is called *open* if it is composed of open sets, and *closed* if it is composed of closed sets.

6.11. *Every two open (closed) coverings Ω^α and Ω^β of a topological space R have a common simple subdivision Ω^γ.*

It is sufficient to take for the elements of the covering Ω^γ the sets $|e_i^\alpha| \cap |e_j^\beta|$, where $e_i^\alpha \in \Omega^\alpha$, $e_j^\beta \in \Omega^\beta$.

6.121. *Let*

$$\Omega_\cdot^\alpha = \{e_i^\alpha\}, \qquad i = 1, 2, \cdots, s_\alpha,$$

be an open covering of a T_1-space R. If at least one of the sets e_i^α contains more than one point, then there exists an open covering $\Omega^\beta = \{e_j^\beta\}$ following Ω^α,

$$\Omega^\beta > \Omega^\alpha.$$

In fact, let, for instance, e_1^α contain at least two points a and a'. Put

$$e_i^\beta = e_i^\alpha - a \text{ for } i \leqq s_\alpha,$$
$$e_i^\beta = e_1^\alpha - a \text{ for } i = s_\alpha + 1.$$

The so-defined covering $\Omega^\beta = \{e_i^\beta\}$, $i=1, 2, \cdots, s_\alpha+1$, is evidently a subdivision of the covering Ω^α. But Ω^α is not a subdivision of Ω^β, since e_1^α is not contained in any of the sets e_i^β.

6.122. *Let R be a T_1-space consisting of an infinite number of points. Then every closed covering*

$$\Omega^\alpha = \{e_i^\alpha\}, \qquad i = 1, 2, \cdots, s_\alpha, \tag{6.1}$$

of R possesses a subdivision Ω^β, containing an element with an infinite number of points not belonging to any other element of the covering Ω^β.

In fact, let us delete from (6.1) one after another all elements, all points of which with the possible exception of a finite number of them are contained in the sum of the following elements of the covering Ω^α. At every such deletion we lose not more than a finite number of points of the space R. Hence, if R consists of an infinite number of points, we shall at last reach such a first element e_i that the set $O_i = e_i - (e_{i+1} \cup \cdots \cup e_{s_\alpha})$ contains infinitely many points. If at the preceding deletion we lost the finite set of points $p_1, \cdots, p_h$, then

$$\{p_1, \cdots, p_h, e_i^\alpha, e_{i+1}^\alpha, \cdots, e_{s_\alpha}^\alpha\}$$

is the required subdivision of the covering Ω^α.

6.123. *Let R be a T_2-space consisting of an infinite number of points. For every closed covering*

$$\Omega^\alpha = \{e_i^\alpha\}, \qquad i = 1, 2, \cdots, s_\alpha, \tag{6.2}$$

of the space R there is a covering

$$\Omega^\beta = \{e_j^\beta\}$$

following after Ω^α.

In fact, we may suppose that Ω^α satisfies the conditions of 6.122 and that, for instance,

$$O_1 = e_1^\alpha - \bigcup_{2 \leq i \leq s_\alpha} e_i^\alpha$$

contains infinitely many points. Take two points a and a' of the set O_1 and choose such a neighbourhood O_a of the point a that

$$\overline{O}_a \subset O_1 - a'.$$

Since O_1 is an open set, such a neighbourhood may be found. Put now

$$\overset{\beta}{e_i} = \overset{\alpha}{e_i} - O_\alpha, \quad \text{if } i \leqq s_\alpha,$$
$$\overset{\beta}{e_i} = \overline{O}_\alpha, \quad \text{if } i = s_\alpha + 1.$$

The covering $\Omega^\beta = \{\overset{\beta}{e_i}\}$, $i = 1, 2, \cdots, s_\alpha + 1$, is the required covering.

From what has been proved above follows

6.12. *The system of all open coverings of every T_1-space consisting of an infinite number of points and the system of all closed coverings of every T_2-space consisting of an infinite number of points are unbounded systems of coverings.*

6.2. From now on and until the end of the present paper we shall suppose, if the contrary is not explicitly stated, that R is a normal space consisting of an infinite number of points. *By* a *covering of the space R we shall always mean an open covering*. The system of all open coverings of the space R we shall denote by $\mathfrak{O}$.

By A we always denote a closed set lying in the space R. A itself is a normal space, which, in general, cannot be asserted with respect to $R-A$.

In every covering $\Omega = \{e_i\}$, $i = 1, 2, \cdots, s$, of the space R we distinguish:

1°. *Elements of the first kind*, i.e., elements meeting A; we denote them by $e_1, \cdots, e_p$.

2°. *Elements of the second kind*, $e_{p+1}, \cdots, e_s$, not meeting A.

The elements of the second kind are subdivided into *boundary elements*:

$$e_{p+1}, \cdots, e_q,$$

satisfying the condition $A \cap \bar{e}_i \neq 0$, $i = p+1, \cdots, q$, and *inner elements*:

$$e_{q+1}, \cdots, e_s,$$

for which $\bar{e}_i \subset R - A$, $i = q+1, \cdots, s$.

DEFINITION 6.21. *A covering*

$$\Omega = \{e_i\}, \qquad i = 1, 2, \cdots, s,$$

of the space R is called regular with respect to A, if it satisfies the following conditions:

1°. *The covering Ω contains no boundary elements of second kind.*

2°. *If for some elements of the first kind $e_{i_0}, \cdots, e_{i_r}$ we have*

$$A \cap e_{i_0} \cap \cdots \cap e_{i_r} = 0,$$

then

$$\bar{e}_{i_0} \cap \cdots \cap \bar{e}_{i_r} = 0.$$

Observe that from these conditions follows

3°. *If for any elements $e_{i_0}, \cdots, e_{i_r}$ of the covering Ω we have*

$$A \cap e_{i_0} \cap \cdots \cap e_{i_r} = 0,$$

then

$$A \cap \bar{e}_{i_0} \cap \cdots \cap \bar{e}_{i_r} = 0.$$

6.22. *Every covering*

$$\Omega^\alpha = \{e_i^\alpha\}, \qquad i = 1, 2, \cdots, s_\alpha,$$

of the space R has a subdivision Ω^β regular with respect to A.

Proof. Consider the open covering

$$A\Omega^\alpha = \{Ae_1^\alpha, \cdots, Ae_p^\alpha] \tag{6.21}$$

of the set A. Since A is a normal space, there exists a closed covering

$$a_1, \cdots, a_p \tag{6.22}$$

of the set A similar to it and inscribed into (6.21). About the system of closed sets (6.22) we circumscribe a system of sets

$$Oa_1, \cdots, Oa_p \tag{6.23}$$

open in R and similar to it such that for $j=1, 2, \cdots, p$

$$a_j \subset Oa_j \subset e_j^\alpha. \tag{6.24}$$

For each a_j take a neighbourhood $O'a_j$ such that

$$\overline{O'a_j} \subset Oa_j \tag{6.25}$$

and put

$$e_j^\beta = O'a_j, \qquad j = 1, 2, \cdots, p. \tag{6.26}$$

Take further a neighbourhood $O''A$ of the set A such that

$$\overline{O''A} \subset \bigcup_{1 \leq j \leq p} e_j^\beta$$

and denote the non-void sets among the $e_i^\alpha - \overline{O''A}$ by e_j^β, $j=p+1, \cdots, s_\beta$.

These last exhaust all elements of the second kind of the covering

$$\Omega^\beta = \{e_j^\beta\}, \qquad j = 1, 2, \cdots, s_\beta,$$

and all these elements are evidently inner elements.

Thus the covering Ω^β satisfies the first condition of regularity with respect to A.

Let us prove that the second condition of regularity is also satisfied. Sup-

pose that for certain elements of the first kind, which we shall for simplicity's sake denote by $e_1^\beta, \cdots, e_r^\beta$, we have

$$A \cap e_1^\beta \cap \cdots \cap e_r^\beta = 0$$

and, consequently, also

$$Ae_1^\beta \cap \cdots \cap Ae_r^\beta = 0.$$

Since $a_i \subset Ae_i^\beta$, we have also $a_1 \cap \cdots \cap a_r = 0$ and, consequently also

$$Oa_1 \cap \cdots \cap Oa_r = 0.$$

And moreover

$$\overline{O'a_1} \cap \cdots \cap \overline{O'a_r} = 0,$$

i.e.,

$$\bar{e}_1^\beta \cap \cdots \cap \bar{e}_r^\beta = 0.$$

Since Ω^β is by very construction a subdivision of Ω^α, (6.22) is proved.

6.3. Consider in any covering Ω^α of the space R the set of elements of the first kind and of boundary elements of the second kind. These elements considered as vertices of the nerve K^α of the covering Ω^α define in K^α a subcomplex K_1^α. By K_0^α we as always understand the nerve of the covering $A\,\Omega^\alpha$ considered as a subcomplex of the complex K^α. Evidently we always have $K_0^\alpha \subset K_1^\alpha$.

6.31. *If a covering Ω^α is regular with respect to A, then $K_1^\alpha = K_0^\alpha$.*

In fact, from the first condition of regularity it follows that the complexes K_1^α and K_0^α have the same vertices, while from the second condition it follows that every simplex of K_1^α is at the same time a simplex of K_0^α.

At every projection of the covering Ω^β into the covering Ω^α the complex K_1^β is obviously transformed into K_1^α; hence we may speak of the spectrum

$$[B_\nabla^r(K^\alpha - K_1^\alpha - C^\alpha); \pi_\beta^\alpha], \tag{6.31}$$

where, as always, C^α denotes the special subcomplex of the complex K^α.

The elements of the spectrum (6.31) corresponding to coverings Ω^α regular with respect to A form in virtue of 6.22 a cofinal part of this spectrum. Hence, having in view 6.31, we obtain

6.32. *The groups*

$$\lim\, [B_\nabla^r(K^\alpha - K_0^\alpha - C^\alpha); \pi_\beta^\alpha],$$
$$\lim\, [B_\nabla^r(K^\alpha - K_1^\alpha - C^\alpha); \pi_\beta^\alpha]$$

are isomorphic.

7. Betti groups of topological spaces (first definition)

7.1. Let R be a normal space consisting of an infinite number of points. Let $\mathfrak{O}=\{\Omega^\alpha\}$ be the system of all open coverings of the space R. By K^α we always denote the nerve of the covering Ω^α.

Definition 7.11. *The group*(9)

$$\lim\ [B^r_\nabla(K^\alpha);\ \pi^\alpha_\beta], \tag{7.1}$$

where π^α_β is a projection of the group $B^r_\nabla(K^\alpha)$ into the group $B^r_\nabla(K^\beta)$, generated by any projection of the covering Ω^β into the covering Ω^α will be denoted by $\mathfrak{B}^r_\nabla(R)$.

7.2. Let G be a topological space consisting of an infinite number of points and homeomorphic to an open set of a certain normal space (in particular, for G may be taken any normal space). By a *special subset of the space G* we shall mean any set $E\subset G$ *whose closure in G is not bicompact.* The system $\mathfrak{C}$ of special subsets of the space G obviously satisfies the condition 5.31: any set $E\subset G$ containing a special subset is itself special. By $\mathfrak{O}=\{\Omega^\alpha\}$ we denote the system of all open coverings of the space G. Special elements of the covering Ω^α, i.e., elements e^α_i, for which $|e^\alpha_i|$ is a special subset, determine the special subcomplex C^α of the complex K^α: the complex $C^\alpha\subset K^\alpha$ consists of simplexes of the complex K^α, all vertices of which are special elements of Ω^α.

Definition 7.2. *The group*(9)

$$\lim\ [B^r_\nabla(K^\alpha - C^\alpha);\ \pi^\alpha_\beta], \tag{7.2}$$

where π^α_β is a projection of the group $B^r_\nabla(K^\alpha - C^\alpha)$ into $B^r_\nabla(K^\beta - C^\beta)$, generated by any projection of the covering Ω^β into the covering Ω^α is called the r-dimensional (inner) Betti ∇-group of the space G and is denoted by $B^r_\nabla(G)$.

7.3. The field of coefficients forming the foundation of the above definitions is, as always in the ∇-theory, supposed to be a discrete commutative group X. If Ξ is the bicompact group dual to X and

$$B^r_\Delta(K^\alpha) = B^r_\Delta(K^\alpha, \Xi),\qquad B^r_\Delta(K^\alpha \bmod C^\alpha) = B^r_\Delta(K^\alpha \bmod C^\alpha, \Xi),$$

then the limit groups

$$B^r_\Delta(R) = \underleftarrow{\lim}\ [B^r_\nabla(K^\alpha);\ \tilde\omega^\beta_\alpha],$$

$$B^r_\Delta(R) = \underleftarrow{\lim}\ [B^r_\Delta(K^\alpha \bmod C^\alpha);\ \tilde\omega^\beta_\alpha]$$

(where $\tilde\omega^\beta_\alpha$ is the homomorphism of the group $B^r_\Delta(K^\beta)$ into $B^r_\Delta(K^\alpha)$, respectively of the group $B^r_\Delta(K^\beta \bmod C^\beta)$ into $B^r_\Delta(K^\alpha \bmod C^\alpha)$, generated by the projection

(9) See 4.8.

of K^β into K^α) are dual to, respectively, the groups $B^r_\nabla(R)$ and $B^r_\nabla(R)$; the group $B^r_\Delta(R)$ is called the r-dimensional (inner) Betti Δ-group of the space R.

Remark. The simple coverings obviously form a cofinal part of the system of all coverings. In the definitions just given we may therefore always assume that all coverings are simple.

7.4. THEOREM 7.41. *Let each of the groups $\mathfrak{B}^r_\nabla(R)$ and $\mathfrak{B}^{r+1}_\nabla(R)$ consist of the zero element only. Let A be a closed set of the space R. Then $\mathfrak{B}^r_\nabla(A)$ is isomorphic to the groups*

$$\lim\,[B^{r+1}_\nabla(K^\alpha - K^\alpha_0);\ \pi^\alpha_\beta),\qquad \lim\,[B^{r+1}_\nabla(K^\alpha - K^\alpha_1);\ \pi^\alpha_\beta],$$

where the complexes K^α_0 and K^α_1 are defined as in 6.3.

Proof. In the first place, for $\mathfrak{C}=0$ in virtue of 6.32 the groups

$$\lim\,[B^{r+1}_\nabla(K^\alpha - K^\alpha_0);\ \pi^\alpha_\beta],\qquad \lim\,[B^{r+1}_\nabla(K^\alpha - K^\alpha_1);\ \pi^\alpha_\beta]$$

are isomorphic, and hence it is sufficient to show that $\mathfrak{B}^r_\nabla(A)$ is isomorphic to the group

$$\lim\,[B^{r+1}_\nabla(K^\alpha - K^\alpha_0);\ \pi^\alpha_\beta].$$

But in virtue of Theorem 4.1 this last group is isomorphic to the group $\mathfrak{B}^r_\nabla(A\mathfrak{O})$. Thus, everything is reduced to the proof of the following

LEMMA 7.411. *The group $\mathfrak{B}^r_\nabla(A)$ is isomorphic to the group $\mathfrak{B}^r_\nabla(A\mathfrak{O})$.*

We begin the proof of Lemma 7.411 with the consideration of the case when A consists of a finite number of points.

Consider coverings Ω^α of the space R satisfying the following conditions:

1°. The covering Ω^α is a simple covering.

2°. Every element of Ω^α contains not more than one point of A.

3°. Two different elements of Ω^α containing points of A do not meet (in particular, no two elements of Ω^α contain one and the same point of A).

It is easily seen that every covering of R has a subdivision satisfying the conditions 1°–3°, so that the system $\mathfrak{B}$ of coverings satisfying these conditions forms a cofinal part of the system $\mathfrak{O}$ of all coverings Ω^α. But if Ω^α satisfies the conditions 1°–3°, then $A\,\Omega^\alpha$ has for its elements the points of A themselves, and the nerve K^α_0 of the covering $A\,\Omega^\alpha$ is a zero-dimensional complex which may be identified with the same finite set A. Hence $B^r_\nabla(K^\alpha_0)=B^r_\nabla(A)$; in the spectrum

$$[B^r_\nabla(K^\alpha_0);\ \pi^\alpha_\beta]$$

the projections π^α_β are identical mappings of $B^r_\nabla(A)$ onto itself and, consequently, the group $B^r_\nabla(A\mathfrak{O})$, being isomorphic to $B^r_\nabla(A\mathfrak{B})$, is also isomorphic to $B^r_\nabla(A)$.

Let now A be infinite. For $\mathfrak{C}_0=0$ in virtue of 5.611 the groups $\mathfrak{B}^r_\nabla(A\mathfrak{D})$ and $\mathfrak{B}^r_\nabla(|A\mathfrak{D}|)$ are isomorphic. Hence it is sufficient to show

7.42. *The group $\mathfrak{B}^r_\nabla(A)$ is isomorphic to the group $B^r_\nabla(|A\mathfrak{D}|)$.*

This proposition follows in its turn from

7.43. *The system of all simple coverings of the space A coincides with the system $|A\mathfrak{D}|$.*

To prove 7.43 it is sufficient to show that every simple open covering

$$\Omega_A = \{a_i\}, \qquad i = 1, 2, \cdots, s,$$

of the set A is an element of the system $|A\mathfrak{D}|$. To this end choose for every $a_i \varepsilon \Omega_A$ a set e_i open in R such that $a_i = A \cap e_i$. The sets e_i in their sum form a certain neighbourhood OA of the set A. Choose a neighbourhood $O'A$ such that $\overline{O'A} \subset OA$ and put

$$e_{s+1} = R - \overline{O'A}.$$

Denoting by $\Omega \varepsilon \mathfrak{D}$ the covering $\{e_i\}$, $i=1, 2, \cdots, s$, of the space R, we evidently have

$$\Omega_A = |A\Omega|,$$

which proves 7.43 and, consequently, 7.42 and 7.41.

7.5. Let now R be a locally bicompact normal space, A an infinite closed subset of the space R. Special subsets of the set R (in particular, of the set A) shall be as above sets, the closures of which are not bicompact. Let

$$(7.51) \qquad \Omega = \{e_1, \cdots, e_s\}$$

be a covering of R. An element $e_i \varepsilon \Omega$ we shall for a moment call unregular if $\bar{e}_i$ is not bicompact and $A \cap \bar{e}_i$ is bicompact and non-void.

7.51. *Every covering* (7.51) *has a subdivision not containing any unregular element.*

For the proof it is sufficient to construct for every covering Ω containing unregular elements such a subdivision Ω_1 that the number of unregular elements in Ω_1 should be by unity less than the number of unregular elements in Ω.

Let e_1 be an unregular element of the covering Ω. Since $\overline{A \cap e_1}$ is bicompact, we may, using the local bicompactness of R, construct such a neighbourhood U_0 of $\overline{A \cap e_1}$ that $\overline{U}_0$ is bicompact. The bicompact set $\overline{U}_0$ may be again enclosed into a neighbourhood U_1 with a bicompact closure. Put

$$e_{11} = e_1 \cap U_1, \qquad e_{12} = e_1 - \overline{U}_0, \qquad \Omega_1 = \{e_{11}, e_{12}, e_2, \cdots, e_s\}.$$

Since e_{11} and e_{12} are regular, Ω_1 is the required subdivision.

7.52. *Let there be given a covering Ω of the space R. As usual denote by K the nerve of the covering Ω, by K_0 the nerve of the covering $A\Omega$, and by C and C_0 special subcomplexes of the complexes K and K_0. If the covering Ω does not contain unregular elements, then $C_0 = K_0 \cap C$.*

In fact, from 7.51 it follows that the complexes C_0 and $K_0 \cap C$ have the same vertices, which correspond in one-to-one manner to special elements of the covering Ω meeting A. Since in C_0 as well as in $K_0 \cap C$ the vertices $e_{i_0}, \cdots, e_{i_r}$ determine a simplex, if

$$A \cap e_{i_0} \cap \cdots \cap e_{i_r} \neq 0,$$

7.52 is proved.

DEFINITION 7.53. *A locally bicompact space R is said to be simply connected with respect to the dimensionality r, if the group $B_\nabla^r(R)$ consists of the zero element only.*

THEOREM 7.54. *Let R be a locally bicompact normal space simply connected with respect to the dimensionalities r and $r+1$. Let A be a closed set of the space R. Then the group $B_\nabla^r(A)$ is isomorphic to the groups*

$$\lim\ [B_\nabla^{r+1}(K^\alpha - K_0^\alpha - C^\alpha);\ \pi_\beta^\alpha], \qquad \lim\ [B_\nabla^{r+1}(K^\alpha - K_1^\alpha - C^\alpha);\ \pi_\beta^\alpha].$$

Proof. Let first A consist of an infinite number of points. In virtue of 7.43 the group $B_\nabla^r(A)$ is isomorphic to the group $B_\nabla^r(|A\mathfrak{O}|, \mathfrak{C}_0)$ (where $\mathfrak{C}_0$ denotes the system of special subsets of the set A) and, consequently, on ground of 5.611, to the group

$$B_\nabla^r(A\mathfrak{O}, \mathfrak{C}_0) = \lim\ [B_\nabla^r(K_0^\alpha - C_0^\alpha);\ \pi_\beta^\alpha].$$

On ground of 7.51 and 7.52 for the cofinal part of the spectrum

$$[B_\nabla^r(K_0^\alpha - C_0^\alpha);\ \pi_\beta^\alpha]$$

(corresponding to the coverings Ω^α not containing unregular elements)

$$K_0^\alpha - C_0^\alpha = K_0^\alpha - C^\alpha$$

so that $\lim\ [B_\nabla^r(K_0^\alpha - C_0^\alpha);\ \pi_\beta^\alpha]$ and consequently also $B_\nabla^r(A)$ are isomorphic to $\lim\ [B_\nabla^r(K_0^\alpha - C^\alpha);\ \pi_\beta^\alpha]$. But this last group is isomorphic in virtue of Theorem 4.1 to the group

$$\lim\ [B_\nabla^r(K^\alpha - K_0^\alpha - C^\alpha);\ \pi_\beta^\alpha]$$

and, consequently, on ground of 6.32, also to the group

$$\lim\ [B_\nabla^r(K^\alpha - K_1^\alpha - C^\alpha);\ \pi_\beta^\alpha],$$

q.e.d.

Let now A be finite. Then $C_0=0$ and, by Lemma 7.411,

$$\overset{r}{B}_\nabla(A) = \overset{r}{B}_\nabla(A) = \overset{r}{B}_\nabla(AO).$$

For coverings Ω^α not containing unregular elements we have

$$\overset{\alpha}{K_0} = \overset{\alpha}{K_0} - \overset{\alpha}{C},$$

so that the group $B^r_\nabla(A)$ is isomorphic to the group

$$\lim\ [\overset{r}{B}_\nabla(\overset{\alpha}{K_0} - \overset{\alpha}{C}); \overset{\alpha}{\pi_\beta}].$$

The rest of the proof is the same as in the case of an infinite A.

7.6. THEOREM 7.6. *Let R be a connected bicompact space simply-connected with respect to the dimensionality* 1. *For any closed set $A \subset R$ the group $B^{00}_\nabla(A)$ is isomorphic to the groups*

$$\lim\ [\overset{1}{B}_\nabla(\overset{\alpha}{K} - \overset{\alpha}{K_0}); \overset{\alpha}{\pi_\beta}], \qquad \lim\ [\overset{1}{B}_\nabla(\overset{\alpha}{K} - \overset{\alpha}{K_1}); \overset{\alpha}{\pi_\beta}].$$

For the proof observe in the first place that from the connectivity of R follows the connectivity of all complexes K^α. Noting this, we construct the proof of Theorem 7.6 precisely on the same lines as the proof of Theorem 7.54 with the only deviation that instead of taking reference to Theorem 4.1 we now apply Theorem 4.6.

8. THE DUALITY LAW OF KOLMOGOROFF

8.1. By the duality law of Kolmogoroff we mean the set of the two following theorems, the proof of which is the object of the present section.

THEOREM 8.11. *Let r be a natural number, R a locally bicompact normal space simply connected with respect to the dimensionalities r and $r+1$. For any closed set $A \subset R$ the groups $B^r_\nabla(A)$ and $B^{r+1}_\nabla(R-A)$ are isomorphic.*

THEOREM 8.12. *Let R be a connected bicompact normal space simply connected with respect to the dimensionality* 1. *For any closed set A the groups $B^{00}_\nabla(A)$ and $B^1_\nabla(R-A)$ are isomorphic.*

8.2. Put $G=R-A$. We begin the proof of Theorems 8.11 and 8.12 by the consideration of the trivial case when G consists of a finite number of points. As regards Theorem 8.12 in this case, from the connectivity of R it follows that the number of (necessarily isolated) points, of which G consists, cannot exceed 1, so that we have either the case when G consists of one point and A is void, or the case when G is void and $A=R$. Since in both these cases Theorem 8.12 is true, it is proved for a finite G.

Let us prove Theorem 8.11 under the assumption of a finite G. In this case $B^{r+1}_\nabla(G)=0$. But for $r>0$ the abstraction from the space R of a finite

number of isolated points does not, as may be easily seen, affect $B^r_\Delta(R)$, so that the group $B^r_\nabla(A) = B^r_\nabla(R-G)$ is isomorphic to the group $B^r_\nabla(R)$, which, by our assumptions, consists of the zero element only and hence is isomorphic to $B^{r+1}_\nabla(G)$. Thus for a finite G Theorem 8.11 is also true.

Let now G consist of an infinite number of points. In virtue of Theorems 7.54 and 7.6 for the proof of Theorems 8.11 and 8.12 it is sufficient to prove the following proposition:

8.21. *For any natural number r and any locally bicompact normal space R the groups $B^r_\nabla(G)$ and* $\lim\ [B^r_\nabla(K^\alpha - K^\alpha_1 - C^\alpha);\ \pi^\alpha_\beta]$ *are isomorphic.*

8.3. **The proof of 8.2.**

DEFINITION 8.31. *Let*

$$\Omega = \{e_1, \cdots, e_p, e_{p+1}, \cdots, e_s\}$$

be a covering of the space R. Denote by ϕ_Ω the sum of all those sets $\bar{e}_i$, which are bicompact and lie in G. The covering Ω is called regular with respect to G, if it satisfies the following two conditions:

1°. *No element of the first kind of the covering Ω meets ϕ_Ω.*

2°. *The elements of the second kind of the covering Ω form a covering of the space G, which we denote by $G\Omega$:*

$$G\Omega = \{e_{p+1}, \cdots, e_s\}.$$

8.32. *Every covering*

$$\Omega = \{e_1, \cdots, e_p, \cdots, e_s\}$$

has a subdivision regular with respect to G.

Proof. Denote all non-void sets of the form $G \cap e_i$ by

$$g_1, \cdots, g_h;$$

they form a covering Γ of the set G. Denote by ϕ_Γ the sum of those among the the sets $\bar{g}_i \subset G$ which are bicompact. After this consider the sets $e_i - \phi_\Gamma$, $i = 1, 2, \cdots, p$, and denote them by

$$e_1', \cdots, e_{p'}'.$$

The covering

$$\Omega' = \{e_1', \cdots, e_{p'}', g_1, \cdots, g_h\}$$

of the space R is a subdivision of the covering Ω. Since $\phi_\Gamma = \phi_{\Omega'}$, Ω' is regular with respect to G.

8.33. *For every covering*

$$\Gamma = \{g_1, \cdots, g_h\}$$

of the set G there is a covering Ω of the space R regular with respect to G and such that

$$G\Omega = \Gamma.$$

In fact, denote by ϕ_Γ the sum of those of the sets $\bar{g}_i$ which are bicompact and lie in G. The covering Ω consisting of all elements of Γ and of the set $R-\phi_\Gamma$ is the required covering.

Remark. Among the elements of the second kind of the covering Ω (regular with respect to G) those and only those are special elements of the covering $G\Omega$ which satisfy at least one of the following two conditions:

1°. They are special elements in Ω.

2°. Their closure meets A.

8.4. Consider the system $\mathfrak{O}_G$ of all coverings

$$\Gamma^\alpha = \{g_k^\alpha\}, \qquad k = 1, 2, \cdots, h_\alpha,$$

of the set G. Denote by

$$\Omega^{\alpha\lambda} = \{e_i^{\alpha\lambda}\} \cup \{g_k^\alpha\}, \qquad i = 1, 2, \cdots, p_{\alpha\lambda}, \ k = 1, 2, \cdots, h_\alpha,$$

every regular with respect to G covering of the space R having Γ^α for the set of its elements of the second kind, i.e., satisfying the condition

$$G\Omega^{\alpha\lambda} = \Gamma^\alpha.$$

For convenience we shall sometimes write instead of g_k^α also $e_i^{\alpha\lambda}$, where $i=p_{\alpha\lambda}+k$. The nerves of $\Omega^{\alpha\lambda}$ and Γ^α we denote respectively by $K^{\alpha\lambda}$ and K^α; the special subcomplexes of $K^{\alpha\lambda}$ and K^α we denote respectively by $C^{\alpha\lambda}$ and C^α. By $K_1^{\alpha\lambda}$ we denote, as usual, the subcomplex of $K^{\alpha\lambda}$ determined by those vertices $e_i^{\alpha\lambda} \in K^{\alpha\lambda}$ for which

$$A \cap \bar{e}_i^{\alpha\lambda} \neq 0.$$

Let us prove the identity

$$K^\alpha - C^\alpha = K^{\alpha\lambda} - K_1^{\alpha\lambda} - C^{\alpha\lambda}. \tag{8.41}$$

Let $t \in K^\alpha - C^\alpha$. The vertices of t are some elements of the second kind $g_k^\alpha \in K^\alpha \in K^{\alpha\lambda}$; suppose they are $g_1^\alpha, g_2^\alpha, \cdots, g_r^\alpha$. We have

$$g_1^\alpha \cap \cdots \cap g_r^\alpha \neq 0.$$

Since $t \notin C^\alpha$, among these g_k^α there is at least one, which is not a special element of the covering Γ^α, i.e., has a bicompact closure lying in G; consequently t can belong neither to $K_1^{\alpha\lambda}$ nor to $C^{\alpha\lambda}$, and so $t \in K^{\alpha\lambda} - K_1^{\alpha\lambda} - C^{\alpha\lambda}$. Conversely, if

$$t \in K^{\alpha\lambda} - K_1^{\alpha\lambda} - C^{\alpha\lambda},$$

then among the vertices of t there is at least one belonging neither to $K_1^{\alpha\lambda}$ nor to $C^{\alpha\lambda}$, i.e., representing a certain $e_i^{\alpha\lambda}$ with a bicompact closure lying in G. In virtue of the regularity of $\Omega^{\alpha\lambda}$ with respect to G all remaining vertices of t are elements of the second kind of the covering $\Omega^{\alpha\lambda}$, i.e., elements of Γ^α, so that $t \in K^\alpha$. Since among the vertices of t at least one has a bicompact closure in G, $t \notin C^\alpha$, and, consequently, $t \in K^\alpha - C^\alpha$.

Since $K^\alpha \subset K^{\alpha\lambda}$, from 8.41 follows

$$C^\alpha \subset K_1^{\alpha\lambda} \cup C^{\alpha\lambda} \tag{8.42}$$

(which, by the way, follows also from the remark made at the end of 8.3).

From (8.41) and (8.42) it follows that the groups $L^r(K^\alpha - C^\alpha)$ and $L^r(K^{\alpha\lambda} - K_1^{\alpha\lambda} - C^{\alpha\lambda})$ are isomorphic: a quite definite isomorphism between these groups is obtained if to every function $f^r \in L^r(K^{\alpha\lambda} - K_1^{\alpha\lambda} - C^{\alpha\lambda})$ is correlated the function $Gf^r \in L^r(K^\alpha - C^\alpha)$, where

$$Gf^r(t^r) = f^r(t^r) \qquad \text{on all } t^r \in K^\alpha.$$

It is easily seen that the isomorphism G is commutative with the operator ∇:

$$\nabla G f^r(t^{r+1}) = G\nabla f^r(t^{r+1}) \tag{8.43}$$

for any $t^{r+1} \in K^\alpha$.

In fact, for any simplex $t^{r+1} \in K^\alpha$ we have

$$\nabla G f^r(t^{r+1}) = \sum_{t^{r+1} > t^r} G f^r(t^r) = \sum_{t^{r+1} > t^r} f^r(t^r) = \nabla f^r(t^r),$$

$$G\nabla f^r(t^{r+1}) = \nabla f^r(t^{r+1}).$$

From (8.43) it follows that the isomorphism G between the groups $L^r(K^{\alpha\lambda} - K_1^{\alpha\lambda} - C^{\alpha\lambda})$ and $L^r(K^\alpha - C^\alpha)$ generates an isomorphic mapping $\phi^{\alpha\lambda}$ of the group $B_\nabla^r(K^{\alpha\lambda} - K_1^{\alpha\lambda} - C^{\alpha\lambda})$ on the group $B_\nabla^r(K^\alpha - C^\alpha)$.

8.5. LEMMA 8.5. *Any two coverings*

$$\Omega^{\alpha\lambda} = \{e_i^{\alpha\lambda}\} \cup \{g_k^\alpha\}, \qquad \Omega^{\beta\mu} = \{e_j^{\beta\mu}\} \cup \{g_l^\beta\}$$

regular with respect to G have such a common subdivision

$$\Omega^{\gamma\nu} = \{e_j^{\gamma\nu}\} \cup \{g_m^\gamma\}$$

regular with respect to G that every element of the second kind g_m^γ of the covering $\Omega^{\gamma\nu}$ is contained in at least one element of the second kind of each of the coverings $\Omega^{\alpha\lambda}$ and $\Omega^{\beta\mu}$.

In fact, let

$$\Omega = \{e_1, \cdots, e_p, g_1, \cdots, g_h\}$$

be any common subdivision of the coverings $\Omega^{\alpha\lambda}$ and $\Omega^{\beta\lambda}$ regular with respect to G. Denote all non-void sets of the form $g_k^\alpha \cap g_l^\beta \cap g_n$ thus:

$$\overset{\gamma}{g}_1, \cdots, \overset{\gamma}{g}_{h_\gamma}$$

Since

$$\bigcup \overset{\alpha}{g}_k = \bigcup \overset{\beta}{g}_l = \bigcup g_n = G,$$

we have also $\bigcup g_m^\gamma = G$. Denote by ϕ_γ the sum of all bicompact $\bar{g}^\gamma \subset G$ and put

$$\overset{\gamma\nu}{e_i} = e_i - \phi_\gamma, \qquad i = 1, 2, \cdots, p.$$

For the covering

$$\Omega^{\gamma\nu} = \{\overset{\gamma\nu}{e_1}, \cdots, \overset{\gamma\nu}{e_p}, \overset{\gamma}{g}_1, \cdots, \overset{\gamma}{g}_{h_\gamma}\}$$

not only $\bigcup_{1 \leq n \leq h_j} g_n^\gamma = G$, but also $\phi_{\Omega^{\gamma\nu}} = \phi_\gamma$; so that $\Omega^{\gamma\nu}$ is regular with respect to G. Besides, for every g_m^γ there are g_k^α, g^β and g_n such that

$$\overset{\gamma}{g}_m = \overset{\alpha}{g}_k \cap \overset{\beta}{g}_l \cap g_n,$$

and consequently

$$\overset{\gamma}{g}_m \subset \overset{\alpha}{g}_k, \qquad \overset{\gamma}{g}_m \subset \overset{\beta}{g}_l.$$

The limit group of the spectrum

$$[B_\nabla^r(K^{\alpha\lambda} - K_1^{\alpha\lambda} - C^{\alpha\lambda}); \overset{\alpha\lambda}{\pi}_{\beta\mu}] \tag{8.51}$$

will not be changed if in this spectrum we retain only elements corresponding to coverings $\Omega^{\alpha\lambda}$ regular with respect to G and put $\Omega^{\beta\lambda} > \Omega^{\alpha\lambda}$ only if $\Omega^{\beta\mu}$ follows after $\Omega^{\alpha\lambda}$ and every element of the second kind of $\Omega^{\beta\mu}$ is contained in some element of the second kind of $\Omega^{\alpha\lambda}$.

Thus, if $B_\nabla^r(K^{\beta\mu} - K_1^{\beta\mu} - C^{\beta\mu}) > B_\nabla^r(K^{\alpha\lambda} - K_1^{\alpha\lambda} - C^{\alpha\lambda})$ in the spectrum (8.51), then

$$B_\nabla^r(K^\beta - C^\beta) > B_\nabla^r(K^\alpha - C^\alpha)$$

in the spectrum

$$[B_\nabla^r(K^\alpha - C^\alpha); \overset{\alpha}{\pi}_\beta]. \tag{8.52}$$

Let us, finally, prove that

$$\overset{\beta\mu}{\phi_\beta} \overset{\alpha\lambda}{\pi}_{\beta\mu} u_{\alpha\lambda} = \overset{\alpha}{\pi}_\beta \overset{\alpha\lambda}{\phi_\alpha} u_{\alpha\lambda} \tag{8.53}$$

for any element $u_{\alpha\lambda}$ of the group $B_\nabla^r(K^{\alpha\lambda} - K_1^{\alpha\lambda} - C^{\alpha\lambda})$. To this end denote by $S_{\alpha\lambda}^{\beta\mu}$ some projection of $\Omega^{\beta\mu}$ into $\Omega^{\alpha\lambda}$ transforming every element of the second kind of $\Omega^{\beta\mu}$ into an element of the second kind of $\Omega^{\alpha\lambda}$. Such a projection $S_{\alpha\lambda}^{\beta\mu}$ generates a projection S_α^β of the covering Γ^β into the covering Γ^α.

We denote the homomorphisms of the groups $L^r(K^{\alpha\lambda}-K_1^{\alpha\lambda}-C^{\alpha\lambda})$ into $L^r(K^{\beta\mu}-K_1^{\beta\mu}-C^{\beta\mu})$ and $L^r(K^\alpha-C^\alpha)$ into $L^r(K^\beta-C^\beta)$ generated by the projections $S_{\alpha\lambda}^{\beta\mu}$ and S_α^β respectively by $\sigma_{\beta\mu}^{\alpha\lambda}$ and σ_β^α.

For the proof of (8.53) it is sufficient to show that for any element $f_{\alpha\lambda}^r \in L^r(K^{\alpha\lambda}-K_1^{\alpha\lambda}-C^{\alpha\lambda})$ and any simplex $t_\beta^r \in K^\beta \subset K^{\beta\mu}$ we have

$$G\overset{\alpha\lambda}{\sigma}_{\beta\mu}\overset{r}{f}_{\alpha\lambda}(\overset{r}{t}_\beta) = \overset{\alpha}{\sigma}_\beta G\overset{r}{f}_{\alpha\lambda}(\overset{r}{t}_\beta). \tag{8.54}$$

The last assertion follows from

$$G\overset{\alpha\lambda}{\sigma}_{\beta\mu}\overset{r}{f}_{\alpha\lambda}(\overset{r}{t}_\beta) = \overset{\alpha\lambda}{\sigma}_{\beta\mu}\overset{r}{f}_{\alpha\lambda}(\overset{r}{t}_\beta) = \overset{r}{f}_{\alpha\lambda}(\overset{\beta\mu}{S}_{\alpha\lambda}\overset{r}{t}_\beta),$$

if we take into account that for $t_\beta^r \in K^\beta$

$$\overset{\beta}{S}_\alpha\overset{r}{t}_\beta = \overset{\beta\mu}{S}_{\alpha\lambda}\overset{r}{t}_\beta.$$

Thus the spectra (8.51) and (8.52) satisfy all conditions of 3.61 and therefore their limit groups

$$\lim\,[\overset{r}{B}_\lambda(\overset{\alpha\lambda}{K} - \overset{\alpha\lambda}{K}_1 - \overset{\alpha\lambda}{C});\, \overset{\alpha\lambda}{\pi}_{\beta\mu}], \qquad \lim\,[\overset{r}{B}_\nabla(\overset{\alpha}{K} - \overset{\alpha}{C});\, \overset{\alpha}{\pi}_\beta]$$

are isomorphic. This proves 8.2 and, consequently, also the duality law of Kolmogoroff.

9. The second definition of Betti groups

9.1. Let $\mathfrak{D}$ be the system of all simple (open) coverings Ω^α of a space R. As always, denote by K^α the nerve of the covering $\Omega^\alpha = \{e_i^\alpha\}$. Special elements of the covering Ω^α we call those sets e_i^α, the closure of which is not bicompact. The nerve of the aggregate of all special elements of the covering Ω^α we denote by C^α and call it the special subcomplex of K^α.

The barycentric subdivision of the complex K^α shall be denoted by $K^{1\alpha}$, the barycentric subdivision of the complex C^α—by $C^{1\alpha}$.

The vertices of the complex $K^{1\alpha}$ are expressions of the form

$$\overset{1\alpha}{e_i} = \overset{\alpha}{e}_{i_0} \cdots \overset{\alpha}{e}_{i_r}, \qquad \overset{\alpha}{e}_{i_k} \in \overset{\alpha}{\Omega},$$

(all $e_{i_k}^\alpha$ entering into $e_i^{1\alpha}$ are different). An aggregate of certain vertices of the complex $K^{1\alpha}$ defines a simplex of this complex if and only if for any two vertices

$$\overset{1\alpha}{e_i} = \overset{\alpha}{e}_{i_0} \cdots \overset{\alpha}{e}_{i_p}, \tag{9.111}$$

$$\overset{1\alpha}{e_j} = \overset{\alpha}{e}_{j_0} \cdots \overset{\alpha}{e}_{j_q} \tag{9.112}$$

of this aggregate all factors in one of the two expressions (9.111) and (9.112), for instance, all factors $e_{j_0}^\alpha, \cdots, e_{j_q}^\alpha$ entering into the expression (9.112), enter also in the other expression, i.e., into (9.111).

It is easily seen that the vertex

$$\overset{1\alpha}{e_i} = \overset{\alpha}{e_{i_0}} \cdots \overset{\alpha}{e_{i_r}}$$

of the complex $K^{1\alpha}$ belongs to the complex $C^{1\alpha}$ (is a *special* vertex of $K^{1\alpha}$) if and only if *all* elements $e^\alpha_{i_0}, \cdots, e^\alpha_{i_r}$ are special elements. We introduce yet the following notation. For any vertex $e^{1\alpha}_i = e^\alpha_{i_0} \cdots e^\alpha_{i_r} \in K^{1\alpha}$ denote by $|e^{1\alpha}_i| = |e^\alpha_{i_0} \cdots e^\alpha_{i_r}|$ the set

$$\overset{\alpha}{e_{i_0}} \cap \cdots \cap \overset{\alpha}{e_{i_r}} \subset R.$$

9.11. *Every projection S^β_α of the complex K^β into K^α generates a simplicial mapping $S^{1\beta}_{1\alpha}$ of the complex $K^{1\beta}$ into the complex $K^{1\alpha}$; moreover, we have*

$$S^{1\beta}_{1\alpha}(C^{1\beta}) \subset C^{1\alpha}.$$

We obtain this mapping correlating to the centre of gravity of any simplex $t_\beta \in K^\beta$ the centre of gravity of its image $S^\beta_\alpha t_\beta$ and observing that $S^\beta_\alpha C^\beta \subset C^\alpha$.

9.2. Definition 9.20. *A covering Ω^α is called multiplicative if the intersection of any number of elements of Ω^α is either void or an element of Ω^α.*

9.21. *Every covering Ω^α has a subdivision Ω^β, which is a multiplicative covering.*

For Ω^β it is sufficient to take the covering consisting of all elements of Ω^α and of all non-void sets, which are intersections of several elements of Ω^α.

Definition 9.22. *By a barycentric subdivision of the covering Ω^α we mean a complex $K^{2\alpha}$, whose vertices are elements e^α_i of the covering Ω^α and whose simplexes are decreasing sequences*

$$\overset{\alpha}{e_{i_0}} \supset \overset{\alpha}{e_{i_1}} \supset \cdots \supset \overset{\alpha}{e_{i_r}}, \qquad e_{i_k} \neq e_{i_{k+1}},$$

of elements of the covering Ω^α.

The complex $K^{2\alpha}$ is evidently a subcomplex of the complex K^α. Put $C^{2\alpha} = K^{2\alpha} \cap C^\alpha$; we call $C^{2\alpha}$ the *special* subcomplex of the complex $K^{2\alpha}$.

Let us come to an agreement, which will enable us to consider $K^{2\alpha}$ also as a subcomplex of the complex $K^{1\alpha}$. To this end observe in the first place the following: if for the vertices of the complex $K^{1\alpha}$

$$\overset{1\alpha}{e_i} = \overset{\alpha}{e_{i_0}} \cdots \overset{\alpha}{e_{i_p}}, \qquad \overset{1\alpha}{e_j} = \overset{\alpha}{e_{j_0}} \cdots \overset{\alpha}{e_{j_q}}$$

we have $|e^{1\alpha}_i| = |e^{1\alpha}_j|$, then, obviously, for $e^{1\alpha}_k$ consisting of all different factors entering into $e^{1\alpha}_i$ or into $e^{1\alpha}_j$ we shall have

$$|\overset{1\alpha}{e_k}| = |\overset{1\alpha}{e_i}| = |\overset{1\alpha}{e_j}|.$$

Therefore, among all expressions

$$\overset{1\alpha}{e_i} = \overset{\alpha}{e_{i_0}} \cdots \overset{\alpha}{e_{i_r}}$$

with one and the same set $|e_i^{1\alpha}|$ we shall have one longest, i.e., one consisting of the maximum number of factors. Correlating to every vertex $e_i^{2\alpha}$ of the complex $K^{2\alpha}$ the longest expression $e_i^{1\alpha}$ satisfying the condition $|e_i^{1\alpha}| = e_i^{2\alpha}$ and observing that from $e_i^{2\alpha} \supset e_j^{2\alpha}$ follows that all factors of $e_j^{1\alpha}$ are contained among the factors of $e_i^{1\alpha}$, we obtain an isomorphic mapping of the complex $K_i^{2\alpha}$ into the complex $K_i^{1\alpha}$. If we now identify $e_i^{2\alpha}$ with $e_i^{1\alpha}$, then we can consider $K^{2\alpha}$ as a subcomplex of the complex $K^{1\alpha}$.

On the other hand, correlating to every vertex $e_i^{1\alpha}$ of the complex $K^{1\alpha}$ the vertex $|e_i^{1\alpha}|$ of the complex $K^{2\alpha}$, we obtain a simplicial mapping $S_{2\alpha}^{1\alpha}$ of the complex $K^{1\alpha}$ on $K^{2\alpha}$, which, if $K^{2\alpha}$ is considered as a subcomplex of the complex $K^{1\alpha}$, leaves all simplexes of the complex $K^{2\alpha}$ fixed. Besides, $S_{2\alpha}^{1\alpha}(C^{1\alpha}) \subset C^{2\alpha}$ and hence:

9.23. *The mapping $S_{2\alpha}^{1\alpha}$ of the complex $K^{1\alpha}$ on the complex $K^{2\alpha}$ generates an isomorphic mapping $\rho_{2\alpha}^{1\alpha}$ of the group $B_\nabla^r(K^{1\alpha}$ mod $C^{1\alpha})$ on the group $B_\nabla^r(K^{2\alpha}$ mod $C^{2\alpha})$ and an isomorphic mapping $\sigma_{1\alpha}^{2\alpha}$ of the group $B_\nabla^r(K^{2\alpha} - C^{2\alpha})$ on the group $B_\nabla^r(K^{1\alpha} - C^{1\alpha})$.*

Correlating to every vertex

$$\overset{1\alpha}{e_i} = \overset{\alpha}{e_{i_0}} \cdots \overset{\alpha}{e_{i_r}} \in \overset{1\alpha}{K}$$

one of the vertices of its bearer in the complex K^α (i.e., one of the $e_{i_k}^\alpha$), we obtain a simplicial mapping $S_\alpha^{1\alpha}$ of the complex $K^{1\alpha}$ on the complex K^α generating, as we know, an isomorphic mapping $\rho_\alpha^{1\alpha}$ of the group $B_\Delta^r(K^{1\alpha}$ mod $C^{1\alpha})$ on the group $B_\Delta^r(K^\alpha$ mod $C^\alpha)$ and, consequently, an isomorphic mapping $\sigma_{1\alpha}^\alpha$ of the group $B_\Delta^r(K^\alpha - C^\alpha)$ on the group $B_\nabla^r(K^{1\alpha} - C^{1\alpha})$.

9.3. Let there be given two multiplicative coverings Ω^α and Ω^β of the space R, of which Ω^β follows after Ω^α. Construct the projection S_α^β of the covering Ω^β into the covering Ω^α in the following way: for every $e_j^\beta \in \Omega^\beta$ we take for $S_\alpha^\beta e_j^\beta$ the *smallest* e_i^α containing e_j^β, i.e., the intersection of all e_i^α containing the given e_j^β. The so-constructed projection is called the *canonical projection* of the covering Ω^β into the multiplicative covering Ω^α.

9.31. *Under the canonical projection S_α^β of the multiplicative covering Ω^β into the multiplicative covering Ω^α the complex $K^{2\beta} \subset K^\beta$ is mapped into $K^{2\alpha} \subset K^\alpha$ so that $C^{2\beta}$ is mapped into $C^{2\alpha}$.*

The second assertion follows from the first, since $C^{2\alpha} = K^{2\alpha} \cap C^\alpha$, $C^{2\beta} = K^{2\beta} \cap C^\beta$ and $S_\alpha^\beta C^\beta \subset C^\alpha$.

For the proof of the first assertion of 9.31 it is sufficient to show

9.311. *If S_α^β is a canonical projection of Ω^β into Ω^α and if $e_i^\beta \supset e_j^\beta$, then $S_\alpha^\beta e_i^\beta \supset S_\alpha^\beta e_j^\beta$.*

In fact,

$$\overset{\beta}{e}_j \subset \overset{\beta}{S}_\alpha \overset{\beta}{e}_j, \qquad \overset{\beta}{e}_j \subset \overset{\beta}{e}_i \subset \overset{\beta}{S}_\alpha \overset{\beta}{e}_i.$$

Since $S^\beta_\alpha e^\beta_j$ is the smallest element of the covering Ω^α containing e^β_j,

$$\overset{\beta}{S}_\alpha \overset{\beta}{e}_i \cap \overset{\beta}{S}_\alpha \overset{\beta}{e}_j = \overset{\beta}{S}_\alpha \overset{\beta}{e}_j,$$

i.e.,

$$\overset{\beta}{S}_\alpha \overset{\beta}{e}_i \supset \overset{\beta}{S}_\alpha \overset{\beta}{e}_j,$$

q.e.d.

Thus every canonical projection S^β_α generates a homomorphic mapping $\bar{\omega}^{2\beta}_{2\alpha}$ of the group $B^r_\Delta(K^{2\beta} \bmod C^{2\beta})$ into the group $B^r_\Delta(K^{2\alpha} \bmod C^{2\alpha})$ and a homomorphic mapping $\pi^{2\alpha}_{2\beta}$ of the group $B^r_\nabla(K^{2\alpha}-C^{2\alpha})$ into the group $B^r_\nabla(K^{2\beta}-C^{2\beta})$.

On ground of 9.11 we know, moreover, that S^β_α generates through the simplicial mapping $S^{\beta 1}_{\alpha 1}$ a homomorphic mapping $\bar{\omega}^{\beta 1}_{\alpha 1}$ of the group $B^r_\Delta(K^{1\beta} \bmod C^{1\beta})$ into the group $B^r_\Delta(K^{1\alpha} \bmod C^{1\alpha})$ and a homomorphic mapping $\pi^{\alpha 1}_{\beta 1}$ of the group $B^r_\nabla(K^{1\alpha}-C^{1\alpha})$ into the group $B^r_\nabla(K^{1\beta}-C^{1\beta})$.

Let us now prove the formulae

$$(9.32)\begin{cases}(9.321) & \bar{\omega}^{1\beta}_{1\alpha} = \left(\rho^{1\alpha}_{\alpha}\right)^{-1} \bar{\omega}^{\beta}_{\alpha}\rho^{1\beta}_{\beta},\\ (9.322) & \bar{\omega}^{2\beta}_{2\alpha} = \rho^{1\alpha}_{2\alpha}\bar{\omega}^{1\beta}_{1\alpha}\left(\rho^{1\beta}_{2\beta}\right)^{-1},\end{cases}$$

$$(9.33)\begin{cases}(9.331) & \pi^{1\alpha}_{1\beta} = \sigma^{\beta}_{1\beta}\pi^{\alpha}_{\beta}\left(\sigma^{\alpha}_{1\alpha}\right)^{-1},\\ (9.332) & \pi^{2\alpha}_{2\beta} = \left(\sigma^{2\beta}_{1\beta}\right)^{-1}\pi^{1\alpha}_{1\beta}\sigma^{2\alpha}_{1\alpha}.\end{cases}$$

On ground of Theorem III (see the Addendum) it is sufficient to prove the formulae (9.32) which may be written in the form

$$(9.34)\begin{cases}(9.341) & \rho^{1\alpha}_{\alpha}\bar{\omega}^{1\beta}_{1\alpha} = \bar{\omega}^{\beta}_{\alpha}\rho^{1\beta}_{\beta},\\ (9.342) & \bar{\omega}^{2\beta}_{2\alpha}\rho^{1\beta}_{2\beta} = \rho^{1\alpha}_{2\alpha}\bar{\omega}^{1\beta}_{1\alpha}.\end{cases}$$

Observe that the homomorphism $\rho^{1\alpha}_{\alpha}\bar{\omega}^{1\beta}_{1\alpha}$ is generated by the simplicial mapping $S^{1\alpha}_{\alpha}S^{1\beta}_{1\alpha}$ of the complex $K^{1\beta}$ into the complex K^{α}; in the same way the homomorphism $\bar{\omega}^{\beta}_{\alpha}\rho^{1\beta}_{\beta}$ is generated by the simplicial mapping $S^{\beta}_{\alpha}S^{1\beta}_{\beta}$ of the complex $K^{1\beta}$ into the complex K^{α}. Hence for the proof of (9.341) it is sufficient to show the truth of the following assertion:

9.32. *For any simplex $t_{1\beta} \in K^{1\beta}$ the simplexes $S^{1\alpha}_{\alpha}S^{1\beta}_{1\alpha}t_{1\beta}$ and $S^{\beta}_{\alpha}S^{1\beta}_{\beta}t_{1\beta}$ are faces of a certain simplex $T_\alpha \in K^\alpha$; moreover, if $t_{1\beta} \in C^{1\beta}$, we may take $T_\alpha \in C^\alpha$.*

In the same way (9.342) follows from

9.33. *For any simplex $t_{1\beta} \in K^{1\beta}$ the simplexes $S^{1\alpha}_{2\alpha}S^{1\beta}_{1\alpha}t_{1\beta}$ and $S^{2\beta}_{2\alpha}S^{1\beta}_{2\beta}t_{1\beta}$ are faces of a certain simplex $T_\alpha \in K^\alpha$; moreover, if $t_{1\beta} \in C^{1\beta}$, we may take $T_\alpha \in C^\alpha$.*

Proof. Let

$$t_{1\beta} = e_0^{1\beta} \cdots e_r^{1\beta}; \qquad e_\mu^{1\beta} = e_{\mu_0}^{\beta} \cdots e_{\mu_q}^{\beta}, \tag{9.35}$$

where $q = q(\mu)$ depends on μ,

$$S_\beta^{1\beta} e_\mu^{1\beta} = e_{\mu_0}^{\beta}, \tag{9.36}$$

$$S_\alpha^{\beta} e_{\mu_0}^{\beta} = e_i^{\alpha} \supset e_{\mu_0}^{\beta} \supset \left| e_\mu^{1\beta} \right|, \tag{9.37}$$

so that

$$S_\alpha^{\beta} S_\beta^{1\beta} e_\mu^{1\beta} \supset \left| e_\mu^{1\beta} \right|. \tag{9.38}$$

If $e_\mu^{1\beta} \varepsilon C^{1\beta}$, then $e_{\mu_0}^{\beta} \varepsilon C^\beta$ and $e_i^\alpha \varepsilon C^\alpha$.

Let

$$S_{1\alpha}^{1\beta} e_\mu^{1\beta} = e_\lambda^{1\alpha} = e_{\lambda_0}^{\alpha} \cdots e_{\lambda_p}^{\alpha}, \qquad S_\alpha^{1\alpha} e_\lambda^{1\alpha} = e_\kappa^{\alpha}.$$

If $e_\mu^{1\beta} \varepsilon C^{1\beta}$, then $e_\lambda^{1\alpha} \varepsilon C^{1\alpha}$, $e_\kappa^\alpha \varepsilon C^\alpha$.

Since

$$e_{\lambda_0}^{\alpha} \supset \left| e_\lambda^{1\alpha} \right| \supset \left| e_\mu^{1\beta} \right|,$$

we have

$$S_\alpha^{1\alpha} S_{1\alpha}^{1\beta} e_\mu^{1\beta} \supset \left| e_\mu^{1\beta} \right|,$$

whence

$$\bigcap_{0 \leq \mu \leq r} S_\alpha^{\beta} S_\beta^{1\beta} e_\mu^{1\beta} \cap S_\alpha^{1\alpha} S_{1\alpha}^{1\beta} e_\mu^{1\beta} \supset \bigcap_{0 \leq \mu \leq r} \left| e_\mu^{1\beta} \right| \neq 0,$$

i.e., all $S_\alpha^\beta S_\beta^{1\beta} e_\mu^{1\beta}$ and $S_\alpha^{1\alpha} S_{1\alpha}^{1\beta} e_\mu^{1\beta}$ are vertices of a certain simplex $T_\alpha \varepsilon K^\alpha$. If all $e_\mu^{1\beta} \varepsilon C^{1\beta}$, then both $S_\alpha^\beta S_\beta^{1\beta} e_\mu^{1\beta}$ and $S_\alpha^{1\alpha} S_{1\alpha}^{1\beta} e_\mu^{1\beta}$ are vertices of C^α and hence $T_\alpha \varepsilon C^\alpha$.

The assertion 9.32 is thus proved.

The assertion 9.33 is proved in the same way. In fact, in the first place we have

$$S_{2\beta}^{1\beta} e_\mu^{1\beta} = \left| e_\mu^{1\beta} \right| = e_\nu^{2\beta}, \qquad S_{2\alpha}^{2\beta} e_\nu^{2\beta} = e_\lambda^{2\alpha} \supset \left| e_\mu^{1\beta} \right|,$$

and if $e_\mu^{1\beta} \varepsilon C^{1\beta}$, then $e_\lambda^{2\alpha} \varepsilon C^{2\alpha}$.

Thus

$$S_{2\alpha}^{2\beta} S_{2\beta}^{1\beta} e_\mu^{1\beta} \supset \left| e_\mu^{1\beta} \right|. \tag{9.391}$$

Let

$$S_{1\alpha}^{1\beta} e_\mu^{1\beta} = e_\kappa^{1\alpha} = e_{\kappa_0}^{\alpha} \cdots e_{\kappa_s}^{\alpha}.$$

Again for $e_\mu^{1\beta} \varepsilon C^{1\beta}$ we have $e_\kappa^{1\alpha} \varepsilon C^{1\alpha}$,

$$S^{1\alpha}_{2\alpha}\overset{1\alpha}{e_\kappa} = \overset{2\alpha}{e_\eta} = |\overset{1\alpha}{e_\kappa}| \supset |\overset{1\beta}{e_\mu}|,$$

$$S^{1\alpha}_{2\alpha}S^{1\beta}_{1\alpha}\overset{1\beta}{e_\mu} \supset |\overset{1\beta}{e_\mu}|, \tag{9.392}$$

and if $\overset{1\beta}{e_\mu} \in C^{1\beta}$, then $e^{2\alpha} \in C^{2\alpha}$.

From the inclusions (9.391) and (9.392) we easily deduce (9.33).

From the formulae (9.32) and (9.33) follow

$$\bar{\omega}^{1\beta}_{1\alpha}\bar{\omega}^{1\gamma}_{1\beta} = \bar{\omega}^{1\gamma}_{1\alpha}, \qquad \bar{\omega}^{2\beta}_{2\alpha}\bar{\omega}^{2\gamma}_{2\beta} = \bar{\omega}^{2\gamma}_{2\alpha}, \tag{9.393}$$

$$\pi^{1\beta}_{1\gamma}\pi^{1\alpha}_{1\beta} = \pi^{1\alpha}_{1\gamma}, \qquad \pi^{2\beta}_{2\gamma}\pi^{2\alpha}_{2\beta} = \pi^{2\alpha}_{2\gamma}. \tag{9.394}$$

Thus we have the inverse and the direct spectra:

$$[B^r_\Delta(K^{1\alpha} \bmod C^{1\alpha}); \bar{\omega}^{1\beta}_{1\alpha}], \qquad [B^r_\Delta(K^{2\alpha} \bmod C^{2\alpha}); \bar{\omega}^{2\beta}_{2\alpha}],$$
$$[B^r_\nabla(K^{1\alpha} - C^{1\alpha}); \pi^{1\alpha}_{1\beta}], \qquad [B^r_\nabla(K^{2\alpha} - C^{2\alpha}); \pi^{2\alpha}_{2\beta}],$$

the limit groups of which are, on the basis of 3.41 and the formulae (9.32), (9.33), isomorphic to the groups $B^r_\Delta(R)$ and $B^r_\nabla(R)$ respectively.

9.4. On the basis of what has been proved we may now formulate the following new definition of the (inner) Betti groups of a space R.

Consider the system $\mathfrak{O}$ of all finite simple multiplicative coverings

$$\Omega^\alpha = \{\overset{\alpha}{e_i}\}, \qquad i = 1, 2, \cdots, s_\alpha,$$

of the space R, whose elements are open sets. Consider the functions

$$\overset{r}{f_\alpha} = \overset{r}{f_\alpha}(\overset{\alpha}{e_{i_0}}, \cdots, \overset{\alpha}{e_{i_r}})$$

with values from a given commutative group J defined for all possible decreasing sequences

$$\overset{\alpha}{e_{i_0}} \supset \overset{\alpha}{e_{i_1}} \supset \cdots \supset \overset{\alpha}{e_{i_r}}$$

consisting of $r+1$ elements of the covering Ω^α. By assumption $f^r_\alpha(e^\alpha_{i_0}, \cdots, e^\alpha_{i_r})$ may be *different from zero only when the sequence*

$$\overset{\alpha}{e_{i_0}} \supset \overset{\alpha}{e_{i_1}} \supset \cdots \supset \overset{\alpha}{e_{i_r}}$$

is strictly decreasing (i.e., when $e^\alpha_{i_\kappa} \neq e^\alpha_{i_{\kappa+1}}$ for $\kappa = 0, 1, \cdots, r-1$). The functions f^r_α form the group $L^r(\Omega^\alpha, J)$. The boundary operators are defined thus:

The lower:

$$\Delta\overset{r}{f_\alpha}(\overset{\alpha}{e_{i_0}}, \cdots, \overset{\alpha}{e_{i_{r-1}}}) = \sum_i (-1)^\kappa \overset{r}{f_\alpha}(\overset{\alpha}{e_{i_0}}, \cdots, \overset{\alpha}{e_{i_{\kappa-1}}}, \overset{\alpha}{e_i}, \overset{\alpha}{e_{i_\kappa}}, \cdots, \overset{\alpha}{e_{i_{r-1}}})$$

(where the summation is extended over all e^α_i for which there exists such an $e^\alpha_{i_\kappa}$ that $e^\alpha_{i_{\kappa-1}} \supset e^\alpha_i \supset e^\alpha_{i_\kappa}$).

The upper:

$$(9.4)\qquad \nabla f_\alpha^r(e_{i_0}^\alpha, \cdots, e_{i_{r+1}}^\alpha) = \sum_\kappa (-1)^\kappa f_\alpha^r(e_{i_0}^\alpha, \cdots, e_{i_{\kappa-1}}^\alpha, e_{i_{\kappa+1}}^\alpha, \cdots, e_{i_{r+1}}^\alpha).$$

(*Remark.* It is easily seen that applying the operators Δ and ∇ to functions satisfying the condition to which the functions f_α^r were just subjected, we obtain functions satisfying the same condition (i.e., vanishing for all sequences $e_{i_0}^\alpha \supset \cdots \supset e_{i_r}^\alpha$ containing equal elements). This assertion is obvious for the operator Δ. Let us prove it for the operator ∇. If in the sequence $e_{i_0}^\alpha \supset \cdots \supset e_{i_{r+1}}^\alpha$ there is more than one pair of coinciding elements, then in each of the sequences $e_{i_0}^\alpha \supset \cdots \supset e_{i_{\kappa-1}}^\alpha \supset e_{i_{\kappa+1}}^\alpha \supset \cdots \supset e_{i_{r+1}}^\alpha$ there is at least one pair of coinciding elements; but then the left-hand side, and consequently the right-hand side of the equality (9.4) vanishes. If, on the other hand, in the sequence $e_{i_0}^\alpha \supset \cdots \supset e_{i_{r+1}}^\alpha$ there is only one pair of coinciding elements, say

$$e_{i_\kappa}^\alpha = e_{i_{\kappa+1}}^\alpha,$$

then on the right-hand side of the equality (9.4) only two terms can be different from zero, namely

$$(-1)^\kappa f_\alpha^r(e_{i_0}^\alpha, \cdots, e_{i_{\kappa-1}}^\alpha, e_{i_{\kappa+1}}^\alpha, e_{i_{\kappa+2}}^\alpha, \cdots, e_{i_{r+1}}^\alpha),$$
$$(-1)^{\kappa+1} f_\alpha^r(e_{i_0}^\alpha, \cdots, e_{i_{\kappa-1}}^\alpha, e_{i_\kappa}^\alpha, e_{i_{\kappa+2}}^\alpha, \cdots, e_{i_{r+1}}^\alpha);$$

but these terms differ only in the sign and their sum is zero, so that again both sides of the equality (9.4) vanish.)

Let, as always, X and Ξ be two groups dual to each other, of which X is discrete and Ξ bicompact. In the group $L^r(\Omega^\alpha, \Xi)$ we consider the following subgroups: The subgroup of *relative cycles* $Z_\Delta^r(\Omega^\alpha, \Xi)$ consisting of all functions f_α^r satisfying the following condition: if $e_{i_0}^\alpha \supset \cdots \supset e_{i_r}^\alpha$ and $\Delta f_\alpha^r(e_{i_0}^\alpha, \cdots, e_{i_r}^\alpha) \neq 0$, then $\bar{e}_{i_r}^\alpha$ is not bicompact. In the group $Z_\Delta^r(\Omega^\alpha, \Xi)$ we consider the subgroup $H_\Delta^r(\Omega^\alpha, \Xi)$ *of relative cycles homologous to zero*: by definition $H_\Delta^r(\Omega^\alpha, \Xi)$ consists of functions f_α^r satisfying the following condition: there exists such a function $f_\alpha^{r+1} \in L^{r+1}(\Omega^\alpha, \Xi)$ that the function $\Delta f_\alpha^{r+1} - f_\alpha^r$ vanishes for all $e_{i_0}^\alpha \supset \cdots \supset e_{i_r}^\alpha$ with bicompact $\bar{e}_{i_r}^\alpha$.

In the group $L^r(\Omega^\alpha, X)$ we consider the subgroup $L_0^r(\Omega^\alpha, X)$ consisting of all functions f_α^r satisfying the following condition: if

$$e_{i_0}^\alpha \supset \cdots \supset e_{i_r}^\alpha$$

and $\bar{e}_{i_r}^\alpha$ is not bicompact, then

$$f_\alpha^r(e_{i_0}^\alpha, \cdots, e_{i_r}^\alpha) = 0.$$

Further we consider the group $Z_\nabla^r(\Omega^\alpha, X)$ of all ∇-cycles contained in $L_0^r(\Omega^\alpha, X)$, i.e., of all functions $f_\alpha^r \in L_0^r(\Omega^\alpha, X)$ satisfying the condition $\nabla f_\alpha^r = 0$. In the group $Z_\nabla^r(\Omega^\alpha, X)$ we consider in its turn the subgroup $H_\nabla^r(\Omega^\alpha, X)$ of

cycles homologous to zero, i.e., the subgroup consisting of functions f_α^r for which there exist such functions $f_\alpha^{r-1} \varepsilon L_0^r(\Omega^\alpha, X)$ that $\Delta f_\alpha^{r-1} = f_\alpha^r$.

(*Remark*. In the case of a bicompact R all these definitions may be simplified: $Z_\Delta^r(\Omega^\alpha, \Xi)$ becomes the group of all cycles, i.e., functions $f_\alpha^r \varepsilon L^r(\Omega^\alpha, \Xi)$ satisfying the condition $\Delta f_\alpha^r = 0$; $H_\Delta^r(\Omega^\alpha, X)$ becomes the subgroup of all cycles homologous to zero, i.e., the subgroup of those functions $f_\alpha^r \varepsilon L^r(\Omega^\alpha, \Xi)$ for which there exist f_α^{r+1} with $\Delta f_\alpha^{r+1} = f_\alpha^r$; the group $L_0^r(\Omega^\alpha, X)$ coincides with $L^r(\Omega^\alpha, X)$; the group $Z_\nabla^r(\Omega^\alpha, X)$ is the group of all ∇-cycles, i.e., of all functions f_α^r with $\nabla f_\alpha^r = 0$; $H_\nabla^r(\Omega^\alpha, X)$ is the group of all ∇-cycles homologous to zero, i.e., of those functions f_α^r for which there exist f_α^{r-1} with $\Delta f_\alpha^{r-1} = f_\alpha^r$.)

The factor groups

$$B_\Delta^r(\Omega^\alpha, J) = Z_\Delta^r(\Omega^\alpha, J) - H_\Delta^r(\Omega^\alpha, J),$$
$$B_\nabla^r(\Omega^\alpha, J) = Z_\nabla^r(\Omega^\alpha, J) - H_\nabla^r(\Omega^\alpha, J)$$

are called, respectively, the Betti Δ- and ∇-groups of the covering Ω^α.

For a bicompact R and $r=0$ we define, besides, the groups $Z_\Delta^{00}(\Omega^\alpha, J)$ and $Z_\nabla^{00}(\Omega^\alpha, J)$. The group $Z_\Delta^{00}(\Omega^\alpha, J)$ is the group of all functions f_α^r, the sum of values of which is equal to zero; the group $Z_\nabla^{00}(\Omega^\alpha, J)$ is the group of all *constant* functions (observe that the group $Z_\nabla^0(\Omega^\alpha, J)$ is the group of all functions constant on every component of the covering Ω^α).

The groups $B_\Delta^{00}(\Omega^\alpha, J)$ and $B_\nabla^{00}(\Omega^\alpha, J)$ we define by the equalities

$$B_\Delta^{00}(\Omega^\alpha, J) = Z_\Delta^{00}(\Omega^\alpha, J) - H_\Delta^0(\Omega^\alpha, J),$$
$$B_\nabla^{00}(\Omega^\alpha, J) = Z_\nabla^0(\Omega^\alpha, J) - Z_\nabla^{00}(\Omega^\alpha, J).$$

If the covering Ω^β is a subdivision of the covering Ω^α, then we make correspond to every element of the covering Ω^β the smallest element of the covering Ω^α containing it. The so-obtained mappings of the covering Ω^β into the covering Ω^α we denote by S_α^β and call them *canonical projections*. To a decreasing sequence of elements of the covering Ω^β under a canonical projection corresponds a decreasing sequence of elements of the covering Ω^α and we have the following homomorphisms:

1°. The homomorphic mapping ρ_α^β of the group $L^r(\Omega^\beta, \Xi)$ into the group $L^r(\Omega^\alpha, \Xi)$ defined by the formula

$$\overset{\beta}{\rho}_\alpha \overset{r}{f}_\beta(\overset{\alpha}{e}_{i_0}, \cdots, \overset{\alpha}{e}_{i_r}) = \sum \overset{r}{f}_\beta(\overset{\beta}{e}_{j_0}, \cdots, \overset{\beta}{e}_{j_r}),$$

where the sum is extended over all $e_{j_0}^\beta \supset \cdots \supset e_{j_r}^\beta$ such that $S_\alpha^\beta e_{j_k}^\beta = e_{i_k}^\alpha$.

2°. The homomorphic mapping σ_β^α of the group $L^r(\Omega^\alpha, X)$ into the group $L^r(\Omega^\beta, X)$ defined by the formula

$$\overset{\alpha}{\sigma}_\beta \overset{r}{f}_\alpha(\overset{\beta}{e}_{j_0}, \cdots, \overset{\beta}{e}_{j_r}) = \overset{r}{f}_\alpha(\overset{\beta}{S}_\alpha \overset{\beta}{e}_{j_0}, \cdots, \overset{\beta}{S}_\alpha \overset{\beta}{e}_{j_r}).$$

The homomorphisms ρ_α^β and σ_β^α preserve respectively the lower and the

upper boundary operators and generate, correspondingly, the homomorphism $\bar{\omega}_\alpha^\beta$ of the group $B_\Delta^r(\Omega^\beta, \Xi)$ into the group $B_\Delta^r(\Omega^\alpha, \Xi)$ and the homomorphism π_β^α of the group $B_\nabla^r(\Omega^\alpha, X)$ into the group $B_\nabla^r(\Omega^\beta, X)$.

If Ξ is, as always, a bicompact topological group, then the groups $B_\Delta^r(\Omega^\alpha, \Xi)$ are also bicompact and we have the inverse spectrum

$$[B_\Delta^r(\Omega^\alpha, \Xi); \bar{\omega}_\alpha^\beta]$$

with a bicompact limit group $B_\Delta^r(R, \Xi)$ called the *r-dimensional Δ-group of the space R to the field of coefficients Ξ.*

Let X be a discrete group; the limit group of the direct spectrum(10)

$$[B_\nabla^r(\Omega^\alpha, X); \pi_\beta^\alpha]$$

is called the *r-dimensional ∇-group of the space R to the field of coefficients X and is denoted by* $B_\nabla^r(R, X)$.

9.41. *If X and Ξ are dual groups and X is discrete and Ξ bicompact, then the groups* $B_\nabla^r(R, X)$ *and* $B_\Delta^r(R, \Xi)$ *are also dual.*

9.42. *In the case of a normal R the just-defined Betti groups remain the same up to an isomorphism, if instead of the system of all open coverings of the space R we consider the system of all closed coverings.*

The proposition 9.41 follows from Theorem 6.1 of Steenrod's paper cited in footnote 4.

The proposition 9.42 may be deduced by the following considerations. The equivalence of the two definitions of Betti groups, namely of the definition given in the present paragraph and the definition given in §7, is proved in the case of open coverings precisely in the same way as in the case of closed coverings; indeed, in the present paragraph we made nowhere use of the fact that the coverings Ω^α consist of open sets. But as regards the equivalence of the definition given in §7 with the analogous definition based on closed coverings, it was proved by Čech(7).

Remark. If the multiplicative coverings Ω^γ, Ω^β, Ω^α follow one after another,

$$\Omega^\gamma > \Omega^\beta > \Omega^\alpha,$$

and S_β^γ, S_α^β, S_α^γ denote the corresponding canonical projections, then the equality

$$S_\alpha^\beta S_\beta^\gamma = S_\alpha^\gamma \tag{9.4}$$

may not be true in spite of the fact that (9.393) and (9.394) always hold.

(10) Here again the definition of the limit-group can be given in a simplified form analogous to that of 4.8.

However, it may be easily shown that the equality (9.4) always holds, if Ω^β is a so-called *exact* subdivision of Ω^α and Ω^γ is an exact subdivision of Ω^β. We say that a subdivision Ω^β of the covering Ω^α is an exact subdivision if it satisfies the following condition: for any $e_i^\alpha \in \Omega^\alpha$, $e_j^\beta \in \Omega^\beta$, the set $e_i^\alpha \cap e_j^\beta$ is an element of the covering Ω^β.

In the first place it is easy to prove that if Ω^γ is an exact subdivision of Ω^β and Ω^β is an exact subdivision of Ω^α, then Ω^γ is an exact subdivision of Ω^α. In fact, let $e_i^\alpha \in \Omega^\alpha$ and $e_k^\gamma \in \Omega^\gamma$ be chosen arbitrarily. Take any $e_j^\beta \supset e_k^\gamma$. Then

$$e_i^\alpha \cap e_k^\gamma = e_i^\alpha \cap e_j^\beta \cap e_k^\gamma = e_h^\beta \cap e_k^\gamma = e_k^\gamma.$$

For the proof of the equality (9.4) in the case of exact subdivisions consider some $e_k^\gamma \in \Omega^\gamma$ and put

$$S_\beta^\gamma e_k^\gamma = e_j^\beta, \qquad S_\alpha^\beta e_j^\beta = e_i^\alpha, \qquad S_\alpha^\gamma e_k^\gamma = e_h^\alpha.$$

Evidently $e_h^\alpha \subset e_i^\alpha$ and hence it is sufficient to show that $e_i^\alpha \subset e_h^\alpha$. To this end consider the set $e_j^\beta \cap e_h^\alpha$. Since Ω^β is an exact subdivision of Ω^α, we have

$$e_j^\beta \cap e_h^\alpha = e_l^\beta \in \Omega^\beta,$$

and (since $S_\beta^\gamma e_k^\gamma = e_j^\beta$, $e_k^\gamma \subset e_l^\beta$)

$$e_j^\beta \cap e_l^\beta = e_j^\beta,$$

i.e.,

$$e_j^\beta \subset e_l^\beta.$$

On the other hand

$$e_l^\beta = e_j^\beta \cap e_h^\alpha \subset e_j^\beta \cap e_i^\alpha = e_j^\beta;$$

hence

$$e_j^\beta = e_l^\beta \subset e_h^\alpha,$$

and consequently $e_i^\alpha \subset e_h^\alpha$, q.e.d.

Observe that any two multiplicative coverings Ω^α and Ω^β have a common exact subdivision Ω^γ. In fact, it is sufficient to take for Ω^γ the covering consisting of all sets of the form $e_i^\alpha \cap e_j^\beta$, where $e_i^\alpha \in \Omega^\alpha$ and $e_j^\beta \in \Omega^\beta$.

10. The connectivity ring

10.1. The definition of the Betti groups given in the preceding paragraph enables us to transfer to these groups the operation of multiplication defined by Alexander([6]). The advantage of the so-obtained theory in comparison with Alexander's theory consists in the independence of our constructions from any arbitrary ordering of the vertices.

Let then be given a commutative ring J (the ring of coefficients), a multiplicative covering

$$\Omega = \{e_i\}, \qquad i = 1, 2, \cdots, s,$$

of the space R and two functions $f^p(e_{i_0}, \cdots, e_{i_p})$ and $f^q(e_{i_0}, \cdots, e_{i_q})$ satisfying the conditions of 9.4. Construct the function $f^{p+q} = [f^p \cdot f^q]$ putting for any decreasing sequence

$$e_{i_0} \supset \cdots \supset e_{i_{p+}}$$

of $p+q+1$ elements of the covering Ω

$$f^{p+q}(e_{i_0}, \cdots, e_{i_{p+q}}) = f^p(e_{i_0}, \cdots, e_{i_p}) f^q(e_{i_p}, \cdots, e_{i_{p+q}}).$$

In the same way as Alexander we deduce by means of simple computation that this multiplication is associative and distributive with respect to addition and that it possesses the following fundamental property:

10.11. $\nabla[f^p \cdot f^q] = [\nabla f^p \cdot f^q] + (-1)^p [f^p \cdot \nabla f^q]$.

From 10.11 immediately follows

10.12. *The product of two cycles is a cycle.*

10.13. *The product of any cycle with a cycle homologous to zero is equal to zero.*

Hence in its turn it follows that the operation of multiplication of functions generates the operation of multiplication of elements of Betti groups: if $z^p \varepsilon B^p_\nabla(\Omega, J)$, $z^q \varepsilon B^q_\nabla(\Omega, J)$, then by $[z^p \cdot z^q]$ we denote the class of homologies $z^{p+q} \varepsilon B^{p+q}(\Omega, J)$ containing the cycle $f^{p+q} = [f^p \cdot f^q]$, where f^p and f^q are arbitrary cycles belonging respectively to the homological classes z^p and z^q.

10.2. Let now be given two elements of the groups $B^p_\nabla(R, J)$ and $B^q_\nabla(R, J)$, i.e., two bundles u^p and u^q of the spectra

$$[B^p_\nabla(\Omega^\alpha, J); \pi^\alpha_\beta], \qquad [B^q_\nabla(\Omega^\alpha, J); \pi^\alpha_\beta].$$

Choose in every bundle an element, $u^p_\alpha \varepsilon u^p, u^q_\alpha \varepsilon u^q$ (with the same α in both cases) and denote by $[u^p \cdot u^q]$ the bundle $u^{p+q} \varepsilon B^{p+q}_\nabla(R, J)$ containing the element $[u^p_\alpha \cdot u^q_\alpha]$.

Let us prove that the so-defined product $[u^p \cdot u^q]$ does not depend on the choice of the elements $u^p_\alpha \varepsilon u^p$ and $u^q_\alpha \varepsilon u^q$. To this end we prove in the first place

10.21. *If $\Omega^\beta > \Omega^\alpha$ and σ^α_β is a mapping of the group $L^r(\Omega^\alpha)$ into the group $L^r(\Omega^\beta)$ generated by some projection S^β_α of the covering Ω^β into the covering Ω^α, then*

$$\sigma^\alpha_\beta[f^p_\alpha \cdot f^q_\alpha] = [\sigma^\alpha_\beta f^p_\alpha \cdot \sigma^\alpha_\beta f^q_\alpha].$$

In fact,

$$\begin{aligned}\overset{\alpha}{\sigma}_\beta[\overset{p}{f}_\alpha\cdot\overset{q}{f}_\alpha](\overset{\beta}{e}_{i_0}\cdots\overset{\beta}{e}_{i_{p+q}}) &= [\overset{p}{f}_\alpha\cdot\overset{q}{f}_\alpha](\overset{\beta}{S}_\alpha\overset{\beta}{e}_{i_0}\cdots\overset{\beta}{S}_\alpha\overset{\beta}{e}_{i_{p+q}})\\ &= \overset{p}{f}_\alpha(\overset{\beta}{S}_\alpha\overset{\beta}{e}_{i_0},\cdots,\overset{\beta}{S}_\alpha\overset{\beta}{e}_{i_p})\overset{q}{f}_\alpha(\overset{\beta}{S}_\alpha\overset{\beta}{e}_{i_p},\cdots,\overset{\beta}{S}_\alpha\overset{\beta}{e}_{i_{p+q}})\\ &= \overset{\alpha}{\sigma}_\beta\overset{p}{f}_\alpha(\overset{\beta}{e}_{i_0},\cdots,\overset{\beta}{e}_{i_p})\overset{\alpha}{\sigma}_\beta\overset{q}{f}_\alpha(\overset{\beta}{e}_{i_p},\cdots,\overset{\beta}{e}_{i_{p+q}})\\ &= [\overset{\alpha}{\sigma}_\beta\overset{p}{f}_\alpha\cdot\overset{\alpha}{\sigma}_\beta\overset{q}{f}_\alpha](\overset{\beta}{e}_{i_0}\cdots\overset{\beta}{e}_{i_{p+q}}).\end{aligned}$$

From 10.21 follows immediately

10.22. $\pi^\alpha_\beta[u^p_\alpha\cdot u^q_\alpha]=[\pi^\alpha_\beta u^p_\alpha\cdot\pi^\alpha_\beta u^q_\alpha]$.

Let now beside $u^p_\alpha \varepsilon u^p$ and $u^q_\alpha \varepsilon u^q$ be chosen $u^p_\beta \varepsilon u^p$, $u^q_\beta \varepsilon u^q$. We shall prove that $[u^p_\alpha\cdot u^q_\alpha]$ and $[u^p_\beta\cdot u^q_\beta]$ belong to one and the same bundle.

From our assumptions follows the existence of such a covering Ω^γ that $\Omega^\gamma>\Omega^\alpha$, $\Omega^\gamma>\Omega^\beta$ and

$$\overset{\alpha}{\pi}_\gamma\overset{p}{u}_\alpha=\overset{\beta}{\pi}_\gamma\overset{p}{u}_\beta,\qquad \overset{\alpha}{\pi}_\gamma\overset{q}{u}_\alpha=\overset{\beta}{\pi}_\gamma\overset{q}{u}_\beta.$$

Hence, on ground of 10.22,

$$\overset{\alpha}{\pi}_\gamma[\overset{p}{u}_\alpha\cdot\overset{q}{u}_\alpha]=[\overset{\alpha}{\pi}_\gamma\overset{p}{u}_\alpha\cdot\overset{\alpha}{\pi}_\gamma\overset{q}{u}_\beta]=[\overset{\beta}{\pi}_\gamma\overset{p}{u}_\beta\cdot\overset{\beta}{\pi}_\gamma\overset{q}{u}_\beta]=\overset{\beta}{\pi}_\gamma[\overset{p}{u}_\beta\cdot\overset{q}{u}_\beta],$$

whence indeed it follows that $[u^p_\alpha\cdot u^q_\alpha]$ and $[u^p_\beta\cdot u^q_\beta]$ belong to one and the same bundle.

DEFINITION 10.23. *The direct sum of the groups* $B^r_\nabla(R, J)$, $r=0, 1, 2, \cdots$, *is denoted by* $B^r_\nabla(R, J)$ *and is called the complete Betti* ∇*-group of the space* R.

From what has been proved it follows that the established operation of multiplication of elements of the group $B_\nabla(R, J)$ transforms this group into a commutative ring. This ring is called the connectivity ring of the space R.

ADDENDUM. ON CERTAIN PROPOSITIONS OF THE THEORY OF GROUPS

1. The theory of characters of commutative groups is taken for granted in the present paper.

If of two groups X and Ξ, of which X is discrete and Ξ is bicompact, one is the group of characters of the other, then the groups X and Ξ are called dual or conjugated. For any $x \varepsilon X$, $\xi \varepsilon \Xi$ we have in this case that

$$\xi(x)=x(\xi)$$

is en element of the group κ denoted by $\xi x=x\xi$. The group of characters of a group G we denote by aG.

2. Let there be given two groups A^α and A^β both discrete and both bicompact; their groups of characters we denote by X^α and X^β. Let there be given a homomorphic mapping ϕ^β_α of the group A^β into the group A^α. To

every element $x_\alpha \varepsilon X^\alpha$ we correlate in the following manner an element $x_\beta = f_\beta^\alpha x_\alpha$ of the group X^β: by definition, the character $x_\beta \varepsilon X^\beta$ of the group A^β maps every element $a_\beta \varepsilon A^\beta$ on the element $x_\alpha \phi_\alpha^\beta a_\beta$ of the group κ. In other words, the character $x_\beta = f_\beta^\alpha x_\alpha$ of the group A^β is determined by the equation

$$(1) \qquad \overset{\alpha}{f_\beta} x_\alpha a_\beta = x_\alpha \overset{\beta}{\phi_\alpha} a_\beta.$$

The mapping ϕ_α^β of the group A^β into the group A^α and the mapping f_β^α of the group X^α into the group X^β are called *conjugated mappings*. The relation of conjugateness of two mappings is a symmetrical relation.

THEOREM I. *Let there be given two isomorphic groups A^α and A^β; denote by ϕ_α^β any isomorphic mapping of A^β on A^α. Then the homomorphism f_β^α conjugated to the isomorphism ϕ_α^β is an isomorphic mapping of the group $X^\alpha = \chi A^\alpha$ on the group $X^\beta = \chi A^\beta$.*

Proof. The mapping f_β^α is defined by the formula (1); in order to prove that f_β^α is an isomorphic mapping it is sufficient to show that for $x_\alpha \varepsilon X^\alpha$, $x_\alpha \neq 0$, we have also $f_\beta^\alpha x_\alpha \neq 0$, i.e., that at least for one $a_\beta \varepsilon A^\beta$

$$\overset{\alpha}{f_\beta} x_\alpha a_\beta \neq 0.$$

Since $x_\alpha \neq 0$, there exists such an $a_\alpha \varepsilon A^\alpha$ that $x_\alpha a_\alpha \neq 0$. Since ϕ_α^β is an isomorphism, there exists such an $a_\beta \neq 0$ (and, moreover, a unique one) that $\phi_\alpha^\beta a_\beta = a_\alpha$. Then

$$\overset{\alpha}{f_\beta} x_\alpha a_\beta = x_\alpha \overset{\beta}{\phi_\alpha} a_\beta = x_\alpha a_\alpha \neq 0,$$

and our assertion is proved.

Thus f_β^α is an isomorphism. Let us prove that f_β^α maps X^α *on* X^β. Let there be given $x_\beta \varepsilon X^\beta$, $x_\beta \neq 0$. We have to find an $x_\alpha \varepsilon X^\alpha$ such that for any $a_\beta \varepsilon A^\beta$

$$\overset{\alpha}{f_\beta} x_\alpha a_\beta = x_\alpha \overset{\beta}{\phi_\alpha} a_\beta = x_\beta a_\beta.$$

Since ϕ_α^β is an isomorphism on A^α, for every $a_\alpha \varepsilon A^\alpha$ there exists a unique $a_\beta \varepsilon A^\beta$ such that $\phi_\alpha^\beta a_\beta = a_\alpha$. Putting

$$x_\alpha a_\alpha = x_\beta a_\beta,$$

we determine the required x_α.

Theorem I is thus proved.

THEOREM II. *Let there be given two groups A^α and A^β respectively dual to the groups X^α and X^β, an isomorphism ϕ_α^β of the group A^β on A^α and the isomorphism ϕ_β^α of the group A^α on A^β dual to the isomorphism ϕ_α^β:*

$$\overset{\alpha}{\phi_\beta} = (\overset{\beta}{\phi_\alpha})^{-1}.$$

Denote by f^α_β *and* f^β_α *the isomorphisms conjugated to the isomorphisms* ϕ^β_α *and* ϕ^α_β. *Then*

$$f^\beta_\alpha = (f^\alpha_\beta)^{-1}.$$

Proof. The isomorphisms f^α_β and f^β_α are determined by the equations

$$f^\alpha_\beta x_\alpha a_\beta = x_\alpha \phi^\beta_\alpha a_\beta, \qquad f^\beta_\alpha x_\beta a_\alpha = x_\beta \phi^\alpha_\beta a_\alpha.$$

Let

$$a_\beta = \phi^\alpha_\beta a_\alpha = (\phi^\beta_\alpha)^{-1} a_\alpha, \qquad x_\beta = f^\alpha_\beta x_\alpha.$$

For any $a_\alpha \varepsilon A^\alpha$ and any $x_\beta \varepsilon f^\alpha_\beta x_\alpha \varepsilon X^\beta$ we have

$$f^\beta_\alpha x_\beta a_\alpha = x_\beta \phi^\alpha_\beta a_\alpha = x_\beta a_\beta = f^\alpha_\beta x_\alpha a_\beta = x^\alpha \phi^\beta_\alpha a_\beta = x_\alpha a_\alpha,$$

$$(f^\alpha_\beta)^{-1} x_\beta a_\beta = (f^\alpha_\beta)^{-1} f^\alpha_\beta x_\alpha a_\alpha = x_\alpha a_\alpha,$$

i.e.,

$$f^\beta_\alpha x_\beta a_\alpha = (f^\alpha_\beta)^{-1} x_\beta a_\alpha,$$

q.e.d.

THEOREM III. *Let the groups*

$$X^\alpha,\ X^\beta,\ X^{1\alpha},\ X^{1\beta}$$

be dual respectively to the groups

$$A^\alpha,\ A^\beta,\ A^{1\alpha},\ A^{1\beta}.$$

Let there be given isomorphic mappings $\rho^{1\alpha}_\alpha$ *and* $\rho^{1\beta}_\beta$ *of* $A^{1\alpha}$ *and* $A^{1\beta}$ *on* A^α *and* A^β *respectively. The conjugated isomorphisms we denote by* $\sigma^\alpha_{1\alpha}$ *and* $\sigma^\beta_{1\beta}$. *Let there be given, besides, homomorphic mappings* $\bar\omega^\beta_\alpha$ *and* $\bar\omega^{1\beta}_{1\alpha}$ *correspondingly of* A^β *into* A^α *and* $A^{1\beta}$ *into* $A^{1\alpha}$, *and the conjugated homomorphisms* π^α_β *and* $\pi^{1\alpha}_{1\beta}$. *Let it be known that*

$$\bar\omega^{1\beta}_{1\alpha} = (\rho^{1\alpha}_\alpha)^{-1} \bar\omega^\beta_\alpha \rho^{1\beta}_\beta. \tag{2}$$

Then

$$\pi^{1\alpha}_{1\beta} = \sigma^\beta_{1\beta} \pi^\alpha_\beta (\sigma^\alpha_{1\alpha})^{-1}.$$

Proof. For the proof it suffices to show that

$$\sigma^\beta_{1\beta} \pi^\alpha_\beta (\sigma^\alpha_{1\alpha})^{-1}$$

satisfies the functional equation

$$\pi^{1\alpha}_{1\beta} x_{1\alpha} a_{1\beta} = x_{1\alpha} \bar\omega^{1\beta}_{1\alpha} a_{1\beta}$$

determining $\pi^{1\alpha}_{1\beta}$, i.e., it is sufficient to prove that for any $x_{1\alpha}$ and $a_{1\beta}$ we have

$$\sigma^{\beta}_{1\beta}\pi^{\alpha}_{\beta}(\sigma^{\alpha}_{1\alpha})^{-1}x_{1\alpha}a_{1\beta} = x_{1\alpha}\bar{\omega}^{1\beta}_{1\alpha}a_{1\beta}. \tag{3}$$

Let us prove this. To this end, replacing $(\sigma^{\alpha}_{1\alpha})^{-1}x_{1\alpha}$ by x_α and $\bar{\omega}^{1\beta}_{1\alpha}$ by its expression (2), we write (3) in the form

$$\sigma^{\beta}_{1\beta}\pi^{\alpha}_{\beta}x_\alpha a_{1\beta} = x_{1\alpha}(\rho^{1\alpha}_{\alpha})^{-1}\bar{\omega}^{\beta}_{\alpha}\rho^{1\beta}_{\beta}\,a_{1\beta}. \tag{4}$$

But $(\rho^{1\alpha}_{\alpha})^{-1}$ and $(\sigma^{\alpha}_{1\alpha})^{-1}$ are conjugated isomorphisms and hence

$$(\sigma^{\alpha}_{1\alpha})^{-1}x_{1\alpha}a_\alpha = x_{1\alpha}(\rho^{1\alpha}_{\alpha})^{-1}a_\alpha.$$

Thus

$$x_{1\alpha}(\rho^{1\alpha}_{\alpha})^{-1}a_\alpha = (\sigma^{\alpha}_{1\alpha})^{-1}x_{1\alpha}a_\alpha = x_\alpha a_\alpha.$$

Substituting this into (4), we obtain as the equality to be proved the following:

$$\sigma^{\beta}_{1\beta}\pi^{\alpha}_{\beta}x_\alpha a_{1\beta} = x_\alpha\bar{\omega}^{\beta}_{\alpha}\rho^{1\beta}_{\beta}\,a_{1\beta}. \tag{5}$$

But

$$\pi^{\alpha}_{\beta}x_\alpha a_\beta = x_\alpha\bar{\omega}^{\beta}_{\alpha}a_\beta, \qquad \pi^{\alpha}_{\beta}x_\alpha = x_\alpha\bar{\omega}^{\beta}_{\alpha}\,\varepsilon\,X^{\beta}$$

and $\sigma^{\beta}_{1\beta}$ and $\rho^{1\beta}_{\beta}$ are conjugated isomorphisms. Therefore the left-hand side of the equality (5) may be transformed to the form

$$\sigma^{\beta}_{1\beta}x_\alpha\bar{\omega}^{\beta}_{\alpha}a_{1\beta} = x_\alpha\bar{\omega}^{\beta}_{\alpha}\rho^{1\beta}_{\beta}\,a_{1\beta},$$

i.e., may be brought to coincidence with the right-hand side of the same equality. The equality (5) and Theorem III are thus proved.

Moscow, U.S.S.R.

PRODUCTS OF NORMAL SEMI-FIELDS

BY

ALBERT NEUHAUS

1. **Introduction.** The normal rings considered by Teichmüller[1] are a generalization of normal fields. They are commutative semi-simple algebras with a group of automorphisms whose order is the order of the algebra. Since they are direct sums of isomorphic fields, Albert[2] called these normal rings semi-fields. In case the group is cyclic he called them cyclic semi-fields.

Albert considered the direct product of two cyclic semi-fields $\mathfrak{Z}$ and $\mathfrak{Y}$ of order n over the reference field. If $[G]$ and $[H]$ are the defining cyclic automorphism groups of $\mathfrak{Z}$ and $\mathfrak{Y}$ respectively he defined the sub-algebra $\mathfrak{W}_k$ of all quantities of $\mathfrak{Z}\times\mathfrak{Y}$ unaltered by $[GH^k]$, for any integer k, and proved

$$\mathfrak{Z}\times\mathfrak{Y} = \mathfrak{Z}\times\mathfrak{W}_k.$$

These factorizations of $\mathfrak{Z}\times\mathfrak{Y}$ may be thought of as being determined by the factorizations

$$[G]\times[H] = [GH^k]\times[H]$$

of the group $[G]\times[H]$.

In the present discussion we shall consider normal semi-fields $\mathfrak{S}$ with an arbitrary group $\mathfrak{G}$. We shall show that we can obtain other direct factorizations of $\mathfrak{S}\times\mathfrak{T}$ with group $\mathfrak{G}\times\mathfrak{G}^*$ if $\mathfrak{G}$ contains a normal divisor $\mathfrak{N}$, such that $\mathfrak{G}/\mathfrak{N}$ is equivalent to a subgroup $\mathfrak{H}_c^*$ of the centrum of $\mathfrak{G}^*$, the group of $\mathfrak{T}$. We shall generalize Albert's cyclic systems to normal systems $(\mathfrak{S}, \mathfrak{G})$. Then we shall see that the set Σ of all normal systems with group $\mathfrak{G}$ with the property

$$\mathfrak{G}/\mathfrak{N}\cong\mathfrak{H}_c$$

($\mathfrak{H}_c$ a subgroup of the centrum of $\mathfrak{G}$) is closed under multiplication, and that this multiplication is associative if $\mathfrak{G}$ is the direct product of $\mathfrak{N}$ and $\mathfrak{H}_c$. Finally Σ is an abelian group if $\mathfrak{G}$ is abelian.

Furthermore as in the case of cyclic systems we shall consider crossed products of normal systems with their groups, and shall derive a connection between the properties of direct products of two such crossed products and the corresponding properties of normal systems.

Presented to the Society, September 8, 1939; received by the editors April 15, 1940. The author wishes to take this opportunity to express his gratitude to Professor A. A. Albert for his helpful criticism and guidance during the writing of this paper.

(1) Deutsche Mathematik, vol. 1 (1936), pp. 92–102, 192–238.

(2) Albert, A. A., Annals of Mathematics, (2), vol. 39 (1938), pp. 669–682, and *Structure of Algebras*, American Mathematical Society Colloquium Publications, vol. 24, 1939.

2. Normal semi-fields. A separable semi-field $\mathfrak{S}$ of order n over the reference field $\mathfrak{K}$ is known([2]) to be the direct sum of t separable, isomorphic fields $\mathfrak{Q}_\mu$ of degree s over $\mathfrak{K}$, that is,

$$\mathfrak{S} = \mathfrak{Q}_1 \oplus \mathfrak{Q}_2 \oplus \cdots \oplus \mathfrak{Q}_t \qquad (st = n).$$

We make the

DEFINITION. *Let $\mathfrak{S}$ be a separable semi-field of order n over $\mathfrak{K}$ and let $\mathfrak{G}=(G_1, G_2, \cdots, G_n)$ be a group of automorphisms of $\mathfrak{S}$. Then we call $\mathfrak{S}$ a normal semi-field with automorphism group $\mathfrak{G}$ if $\mathfrak{S}_\mathfrak{L}=(e^{G_1}, e^{G_2}, \cdots, e^{G_n})$ over $\mathfrak{L}$ for a field $\mathfrak{L}$ over $\mathfrak{K}$ and pairwise orthogonal idempotents e^{G_ν}.*

This is exactly the definition of a normal ring as given by Teichmüller([3]). Diagonal algebras are trivially normal semi-fields and normal fields also are normal semi-fields.

A normal semi-field $\mathfrak{S}$ with group $\mathfrak{G}$ is $\mathfrak{G}$-irreducible([4]), for the diagonal algebra $\mathfrak{S}_\mathfrak{L}$ is irreducible since $\mathfrak{G}$ is transitive with respect to the e^{G_ν}. From Albert's lemma([4]) it follows that $\mathfrak{S}=\mathfrak{Q}\times\mathfrak{E}$ with $\mathfrak{E}=(e^{g_1}, e^{g_2}, \cdots, e^{g_t})$ for a set of g_μ of $\mathfrak{G}$. Teichmüller proved([5]) that $\mathfrak{Q}$ is a normal field over $\mathfrak{K}$, whose automorphism group is the subgroup $\mathfrak{H}$ of $\mathfrak{G}$ which leaves the elements of $\mathfrak{E}$ unaltered.

We prove a result which, for the case where $\mathfrak{S}$ is a field, is a part of the well-known fundamental theorem of the Galois theory([6]).

THEOREM 1. *To every normal divisor $\mathfrak{N}$ of $\mathfrak{G}$ corresponds a normal sub-semi-field $\mathfrak{S}(\mathfrak{N})$ of $\mathfrak{S}$ consisting of all elements of $\mathfrak{S}$ unaltered by the automorphisms of $\mathfrak{N}$, and having $\mathfrak{G}/\mathfrak{N}$ as a group. The order of $\mathfrak{S}(\mathfrak{N})$ is the order of $\mathfrak{G}/\mathfrak{N}$.*

For $\mathfrak{S}(\mathfrak{N})$ is a sub-algebra of a commutative semi-simple algebra and is therefore also commutative and semi-simple, hence $\mathfrak{S}(\mathfrak{N})$ is a semi-field. The group $\mathfrak{G}/\mathfrak{N}$ is a group of automorphisms of $\mathfrak{S}(\mathfrak{N})$ since all elements of a coset of $\mathfrak{N}$ induce the same automorphism in $\mathfrak{S}(\mathfrak{N})$. It is normal since the scalar extension $\mathfrak{S}(\mathfrak{N})_\mathfrak{L}$ is the diagonal sub-algebra of $\mathfrak{S}_\mathfrak{L}$ unaltered by $\mathfrak{N}$ and $\mathfrak{S}(\mathfrak{N})_\mathfrak{L}$ has $(e^{n_1}+e^{n_2}+\cdots+e^{n_i})^{g_\mu}$ $(\mu=1, \cdots, j)$, $ij=n$, as a basis, with $\mathfrak{N}=(n_1, n_2, \cdots, n_i)$, $\mathfrak{G}=g_1\mathfrak{N}+g_2\mathfrak{N}+\cdots+g_j\mathfrak{N}$.

Teichmüller proved this theorem in a different, more difficult way([7]).

The converse, that to every normal semi-field of order j of $\mathfrak{S}$ corresponds a normal divisor of order i of $\mathfrak{G}$, does not hold in general as the following example shows. Let $\mathfrak{S}$ be diagonal with $\mathfrak{G}$, the non-abelian group of order six, as an automorphism group. Then $\mathfrak{G}$ consists of the elements i, s, t, t^2, st, st^2

([3]) Deuring, M., *Algebren*, 1935, p. 96.
([4]) Albert, A. A., *Structure of Algebras*, p. 78.
([5]) Teichmüller, p. 98.
([6]) Albert, A. A., *Modern Higher Algebra*, 1937, p. 174.
([7]) Teichmüller, p. 203.

where i is the identity, s and t of orders two and three respectively, and $ts=st^2$. Then $\mathfrak{S}$ has a normal basis $e^i, e^s, e^t, e^{t^2}, e^{st}, e^{st^2}$ of orthogonal idempotents. The diagonal algebra $\mathfrak{T}$ with (e^i+e^s), (e^t+e^{st}), $(e^{t^2}+e^{st^2})$ as basis is a normal sub-semi-field with (i, t, t^2) as automorphism group. But $\mathfrak{G}$ does not have a normal divisor of order two.

THEOREM 2. *If $\mathfrak{G}$ is the direct product of two subgroups $\mathfrak{N}$, $\mathfrak{H}$, then $\mathfrak{S}$ is the direct product of their corresponding normal sub-semi-fields.*

For, if $\mathfrak{G}=\mathfrak{N}\times\mathfrak{H}$ the $\mathfrak{N}$ and $\mathfrak{H}$ are normal divisors of $\mathfrak{G}$. If n, i, j are the orders of $\mathfrak{G}$, $\mathfrak{N}$, $\mathfrak{H}$ respectively, then $n=ij$. Now $\mathfrak{S}(\mathfrak{N})$ has order j, $\mathfrak{S}(\mathfrak{H})$ order i, hence $\mathfrak{S}(\mathfrak{N})\mathfrak{S}(\mathfrak{H})$ has order less than n if and only if $\mathfrak{S}(\mathfrak{N})$ and $\mathfrak{S}(\mathfrak{H})$ have elements other than those of the reference field $\mathfrak{K}$ in common. But such elements are elements of $\mathfrak{S}$ unaltered by all automorphisms of $\mathfrak{N}$ and $\mathfrak{H}$, hence of $\mathfrak{G}$, and they lie in $\mathfrak{K}$; $\mathfrak{S}(\mathfrak{N})\mathfrak{S}(\mathfrak{H})$ has order n and is equal to $\mathfrak{S}(\mathfrak{N})\times\mathfrak{S}(\mathfrak{H})=\mathfrak{S}$.

3. **Factorizations of direct products of groups.** As we indicated in the introduction, Albert considered direct factorizations of direct products of cyclic semi-fields determined by factorizations of the direct products of their groups. In our later considerations of the direct products we shall generalize the cyclic case by considering subsets of direct products of normal semi-fields unaltered by a subgroup $\mathfrak{G}_0$ of $\mathfrak{G}\times\mathfrak{G}^*$, the direct product of their groups such that $\mathfrak{G}\times\mathfrak{G}^*=\mathfrak{G}_0\times\mathfrak{G}^*$. Before doing so we wish to know about the possibility of such factorizations of the direct products of two groups. The required result is given by

THEOREM 3. *Let $\mathfrak{G}$ and $\mathfrak{G}^*$ be any two finite groups. Then there exist factorizations $\mathfrak{G}\times\mathfrak{G}^*=\mathfrak{G}_0\times\mathfrak{G}^*$ for a subgroup $\mathfrak{G}_0\neq\mathfrak{G}$ of $\mathfrak{G}\times\mathfrak{G}^*$ if and only if there exists a normal divisor $\mathfrak{N}$ of $\mathfrak{G}$ such that $\mathfrak{G}/\mathfrak{N}$ is equivalent to a subgroup $\mathfrak{H}_c^*$ of the centrum of $\mathfrak{G}^*$. The group $\mathfrak{G}_0$ is equivalent to $\mathfrak{G}$ and is contained in $\mathfrak{G}\mathfrak{H}_c^*$.*

For, since $\mathfrak{G}_0$ is a subgroup of $\mathfrak{G}\times\mathfrak{G}^*$ its elements must be products of elements in $\mathfrak{G}$ with elements in $\mathfrak{G}^*$,

$$g_0 = gh^* \qquad (g_0 \text{ in } \mathfrak{G}_0,\ g \text{ in } \mathfrak{G},\ h^* \text{ in } \mathfrak{G}^*);$$

on the other hand every element g_0 must be commutative with every element g^* of $\mathfrak{G}^*$,

$$g^*g_0 = g_0g^*$$

and

$$g^*gh^* = gh^*g^*.$$

Since the elements of $\mathfrak{G}$ are commutative with those of $\mathfrak{G}^*$, we have

$$gg^*h^* = gh^*g^*,$$

hence

$$g^*h^* = h^*g^* \qquad \text{for every } g^* \text{ of } \mathfrak{G}^*.$$

It follows that h^* is in the centrum of $\mathfrak{G}^*$.

Let g_ν $(\nu=1, 2, \cdots, m)$ be the elements of $\mathfrak{G}$, and c_μ^* be elements of the centrum of $\mathfrak{G}^*$. Then we know that every g_0 of $\mathfrak{G}_0$ is uniquely expressible as the product of an element g of $\mathfrak{G}$ and an element c^* of the centrum of $\mathfrak{G}^*$, so that

$$g_0 \equiv gc_g^* = c_g^*g.$$

It follows that $\mathfrak{G}_0$ is equivalent to $\mathfrak{G}$ under the correspondence $g_0 \leftrightarrow g$. Since $\mathfrak{G}_0$ is a group, we have then

$$g_0h_0 = ghc_g^*c_h^* = (gh)_0 = (gh)c_{gh}^*.$$

This holds if and only if $c_g^*c_h^* = c_{gh}^*$. Now it follows that there must exist a homomorphism

$$g \rightarrow c_g^*.$$

Hence these c_g^* must form a group $\mathfrak{H}_c^*$ which, since the c_g^* are in the centrum of $\mathfrak{G}^*$, is a subgroup of this centrum. Thus $\mathfrak{G}$ is homomorphic to a subgroup $\mathfrak{H}_c^*$ of the centrum of $\mathfrak{G}^*$. Consequently(8), there exists a normal divisor $\mathfrak{N}$ of $\mathfrak{G}$, such that $\mathfrak{G}/\mathfrak{N}$ is equivalent to $\mathfrak{H}_c^*$, $\mathfrak{G}/\mathfrak{N} \cong \mathfrak{H}_c^*$.

Conversely, if $\mathfrak{G}$ has a normal divisor $\mathfrak{N}$ such that $\mathfrak{G}/\mathfrak{N}$ is equivalent to a subgroup $\mathfrak{H}_c^*$ of the centrum of $\mathfrak{G}^*$, that is, if $\mathfrak{G} = g_1\mathfrak{N} + g_2\mathfrak{N} + \cdots + g_j\mathfrak{N}$ and $\mathfrak{H}_c^* = (c_1^*, c_2^*, \cdots, c_j^*)$, there is a one-one correspondence $g\mathfrak{N} \leftrightarrow c_g^*$ and hence a homomorphism between $\mathfrak{G}$ and $\mathfrak{H}_c^*$, that is, if $\mathfrak{N} = (n_1, \cdots, n_i)$, $ij = m$, $n_\nu g_\mu \rightarrow c_\mu^*$. We define

$$g_{0\nu,\mu} = n_\nu g_\mu c_\mu^* \qquad (\nu = 1, \cdots, i;\ \mu = 1, \cdots, j).$$

These g_0 form a group $\mathfrak{G}_0$. For, since $\mathfrak{N}$ is a normal divisor of the group $\mathfrak{G}$ we have

$$gn' = n''g, \qquad g_0h_0 = ngc_g^*n'hc_h^* = nn''ghc_g^*c_h^*$$

and

$$nn''gh \rightarrow c_g^*c_h^*.$$

But $\mathfrak{G}/\mathfrak{N}$ is a group and there exists in $\mathfrak{N}$ an element n' such that

$$ngn'h = n_1g_1 \qquad \text{(the unity of } \mathfrak{G}\text{)},$$

$$g_0^{-1} = (ngc_g^*)^{-1} = n'hc_h^* = h_0.$$

The elements of $\mathfrak{G}_0$ are commutative with the elements of $\mathfrak{G}^*$, and $\mathfrak{G}_0$ and $\mathfrak{G}^*$

(8) van der Waerden, B. L. *Moderne Algebra*, vol. 1, 1930, p. 35.

have only the unity element in common, hence

$$\mathfrak{G} \times \mathfrak{G}^* = \mathfrak{G}_0 \times \mathfrak{G}^*$$

and $\mathfrak{G}_0 \leqq \mathfrak{G}\mathfrak{H}_c^*$. This proves our theorem.

4. **Direct products.** In our discussion of direct products of normal semi-fields we shall use

THEOREM 4. *A normal semi-field $\mathfrak{S}$ with group $\mathfrak{G}=(G_\mu; \mu=1, \cdots, n)$ has a normal basis*

$$\mathfrak{S} = (u^{G_1}, u^{G_2}, \cdots, u^{G_n})$$

over $\mathfrak{K}$.

For $\mathfrak{S}=\mathfrak{Q}\times\mathfrak{E}$ with $\mathfrak{Q}$ a normal field of degree s over $\mathfrak{K}$ and $\mathfrak{E}$ a diagonal algebra of order t over $\mathfrak{K}$, $st=n$. It is well-known([9]) that $\mathfrak{Q}$ has a normal basis

$$\xi^{h_1}, \xi^{h_2}, \cdots, \xi^{h_s}$$

with ξ in $\mathfrak{Q}$ and the h_μ the elements of the automorphism group $\mathfrak{H}$ of $\mathfrak{Q}$. On the other hand $\mathfrak{H}$ is the subgroup of $\mathfrak{G}$ whose elements leave the elements of $\mathfrak{E}$ unaltered as we noted in §2. There we saw that $\mathfrak{E}$ has a basis

$$e^{g_1}, e^{g_2}, \cdots, e^{g_t}$$

of orthogonal idempotents, with e in $\mathfrak{E}$ and the g_ν a set of t elements of $\mathfrak{G}$. Hence the n elements $e^{g_\mu}\xi^{h_\nu}$ $(\mu=1, \cdots, t; \nu=1, \cdots, s)$ of $\mathfrak{S}$ form a basis of $\mathfrak{S}$.

Define $u=e^{g_1}\xi^{h_1}$. Then for any element G_σ of $\mathfrak{G}$ the quantity u^{G_σ} is a certain $e^g\xi^h$ in $e^g\mathfrak{Q}$. By taking all n different G of $\mathfrak{G}$ we get all n elements $e^g\xi^h$. Thus $u=e^{g_1}\xi^{h_1}$ and its conjugates form a normal basis of $\mathfrak{S}$.

We now let $\mathfrak{J}=\mathfrak{S}\times\mathfrak{T}$ where $\mathfrak{S}$ and $\mathfrak{T}$ are normal semi-fields of order m and n, and with groups $\mathfrak{G}$ and $\mathfrak{G}^*$ respectively, and prove

THEOREM 5. *The direct product $\mathfrak{J}=\mathfrak{S}\times\mathfrak{T}$ is a normal semi-field of order mn with group $\mathfrak{G}\times\mathfrak{G}^*$.*

For, since $\mathfrak{S}$ and $\mathfrak{T}$ are normal semi-fields there exist two fields $\mathfrak{L}'$ and $\mathfrak{L}''$ over $\mathfrak{K}$, such that $\mathfrak{S}_{\mathfrak{L}'}=(e^{G_1}, e^{G_2}, \cdots, e^{G_m})$ and $\mathfrak{T}_{\mathfrak{L}''}=(f^{G_1^*}, f^{G_2^*}, \cdots, f^{G_n^*})$ over $\mathfrak{L}'$ and $\mathfrak{L}''$ respectively. Let $\mathfrak{L}$ be a composite of $\mathfrak{L}'$ and $\mathfrak{L}''$ over $\mathfrak{K}$. Then since $\mathfrak{L}$ contains $\mathfrak{L}'$ and a field equivalent to $\mathfrak{L}''$ we have

$$\mathfrak{S}_\mathfrak{L} = (e^{G_1}, e^{G_2}, \cdots, e^{G_m}),\ \mathfrak{T}_\mathfrak{L} = (f^{G_1^*}, f^{G_2^*}, \cdots, f^{G_n^*})$$

and $\mathfrak{J}_\mathfrak{L}=\mathfrak{S}_\mathfrak{L}\times\mathfrak{T}_\mathfrak{L}=(e^{G_1}f^{G_1^*}, \cdots, e^{G_1}f^{G_n^*}, e^{G_2}f^{G_1^*}, \cdots, e^{G_m}f^{G_n^*})$ over $\mathfrak{L}$. But $G_\mu G_\nu^*$ $(\mu=1, \cdots, m; \nu=1, \cdots, n)$ are the elements of $\mathfrak{G}\times\mathfrak{G}^*$ and hence $\mathfrak{J}=\mathfrak{S}\times\mathfrak{T}$ is a normal semi-field with $\mathfrak{G}\times\mathfrak{G}^*$ as automorphism group.

Let us now assume that $\mathfrak{G}$ has a normal divisor $\mathfrak{N}$ such that $\mathfrak{G}/\mathfrak{N}\cong\mathfrak{H}_c^*$, a subgroup of the centrum of $\mathfrak{G}^*$. We know from Theorem 2 that $\mathfrak{G}\times\mathfrak{G}^*$

([9]) Deuring, M., Mathematische Annalen, vol. 107 (1932), pp. 140–144.

$=\mathfrak{G}_0\times\mathfrak{G}^*$ where $\mathfrak{G}_0$ has the structure given in Theorem 2. Since the group of $\mathfrak{S}\times\mathfrak{T}$ is $\mathfrak{G}\times\mathfrak{G}^*=\mathfrak{G}_0\times\mathfrak{G}^*$, $\mathfrak{G}_0$ is a normal divisor of $\mathfrak{G}\times\mathfrak{G}^*$. Hence from Theorem 1 it follows that $\mathfrak{J}=\mathfrak{S}\times\mathfrak{T}$ contains a normal sub-semi-field $\mathfrak{W}$ of order n whose elements are unaltered by $\mathfrak{G}_0$ and whose group is $\mathfrak{G}^*$. We prove

THEOREM 6. *The direct product* $\mathfrak{J}=\mathfrak{S}\times\mathfrak{T}=\mathfrak{S}\times\mathfrak{W}$.

For by Theorem 2 $\mathfrak{J}$ is the direct product of $\mathfrak{J}(\mathfrak{G}^*)$ and $\mathfrak{J}(\mathfrak{G}_0)$ since $\mathfrak{G}\times\mathfrak{G}^*=\mathfrak{G}_0\times\mathfrak{G}^*$ and $\mathfrak{J}(\mathfrak{G}^*)=\mathfrak{S}$ and $\mathfrak{J}(\mathfrak{G}_0)=\mathfrak{W}$. We also see that $\mathfrak{G}_0$ can be considered as the group of $\mathfrak{S}$.

For later use we shall construct a basis of $\mathfrak{W}$. Let $\mathfrak{G}=(G_1, G_2, \cdots, G_m)$ have the normal divisor $\mathfrak{N}=(n_1, n_2, \cdots, n_i)$ consisting of i elements of $\mathfrak{G}$ and let $\mathfrak{G}=g_1\mathfrak{N}+g_2\mathfrak{N}+\cdots+g_j\mathfrak{N}$, $ij=m$, $\mathfrak{G}^*=(G_1^*, G_2^*, \cdots, G_n^*)$ with $\mathfrak{H}_c^*=(c_1^*, c_2^*, \cdots, c_j^*)$ a subgroup of the centrum of $\mathfrak{G}^*$, and $\mathfrak{G}/\mathfrak{N}=(g_1\mathfrak{N}, g_2\mathfrak{N}, \cdots, g_j\mathfrak{N})\cong\mathfrak{H}_c^*$. Then every element of $\mathfrak{G}$ is of the form $n_\nu g_\mu$ and every element of $\mathfrak{G}^*$ is a product of a G^* and a c_μ^*. The group $\mathfrak{G}_0$ consists of elements $\mathfrak{G}_{0\nu,\mu}=n_\nu g_\mu c_\mu^*$ ($\nu=1, \cdots, i$; $\mu=1, \cdots, j$). For brevity we write $G_0=ngc_g^*$, where ng is in $g\mathfrak{N}$ and c_g^* is the element of $\mathfrak{H}_c^*$ corresponding to $g\mathfrak{N}$. Let u and v generate normal bases of $\mathfrak{S}$ and $\mathfrak{T}$ respectively. Then the n elements

$$w^{G^*}=\sum_g\left(u^{n_1}+u^{n_2}+\cdots+u^{n_i}\right)^g v^{c_g^* G^*}, \tag{1}$$

where g runs over a system of representatives of $\mathfrak{G}/\mathfrak{N}$, form a normal basis of $\mathfrak{W}$ for all G^* of $\mathfrak{G}^*$. For every w^{G^*} is unaltered by any $H_0=hc_h^*$ of $\mathfrak{G}_0$ if w is. But

$$\begin{aligned} w^{H_0} &= \sum_g\left(u^{n_1}+u^{n_2}+\cdots+u^{n_i}\right)^{gnh} v^{c_g^* c_h^*} \\ &= \sum_g\left(u^{n_1}+u^{n_2}+\cdots+u^{n_i}\right)^{gh} v^{c_g^* c_h^*} \end{aligned}$$

since $gn=n'g$ and $(u^{n_1}+u^{n_2}+\cdots+u^{n_i})^{n'}=(u^{n_1}+u^{n_2}+\cdots+u^{n_i})$. Now gh corresponds to $c_g^*c_h^*$ and $c_g^*c_h^*$ gives all j elements of $\mathfrak{H}_c^*$ if c_g^* runs over all j elements of $\mathfrak{H}_c^*$. Hence

$$w^{G_0}=w, \qquad w^{G^*} \text{ in } \mathfrak{W}.$$

The n elements w^{G^*} are linearly independent. For the linear combination

$$\sum_{G^*} k_{G^*} w^{G^*}, \qquad k_{G^*} \text{ in } \mathfrak{K}, \tag{2}$$

is a linear combination of the mn elements $u^G v^{H^*}$ in $\mathfrak{J}$ with the k_{G^*} as coefficients, since these elements form a basis of $\mathfrak{J}$, (2) can only be zero if all $k_{G^*}=0$. Thus the n elements w^{G^*} of $\mathfrak{W}$ are linearly independent and since $\mathfrak{W}$ has order n they form a normal basis of $\mathfrak{W}$.

Now $\mathfrak{G}/\mathfrak{N}$ and $\mathfrak{H}_c^*$ are abelian groups. If $g\mathfrak{N}\leftrightarrow c_g^*$ is an isomorphism between $\mathfrak{G}/\mathfrak{N}$ and $\mathfrak{H}_c^*$, so is $g\mathfrak{N}\leftrightarrow(c_g^*)^\sigma$ for an arbitrary integer σ. For if

$g\mathfrak{N} \leftrightarrow (c_g^*)^\sigma$, $g\mathfrak{N}h\mathfrak{N} = gh\mathfrak{N} \leftrightarrow (c_g^*)^\sigma (c_h^*)^\sigma = (c_g^* c_h^*)^\sigma$. Each of these isomorphisms yields a $\mathfrak{G}_{0\sigma}$ and a factorization $\mathfrak{G} \times \mathfrak{G}^* = \mathfrak{G}_{0\sigma} \times \mathfrak{G}^*$. Consequently $\mathfrak{J} = \mathfrak{S} \times \mathfrak{T} = \mathfrak{S} \times \mathfrak{W}_\sigma$, where $\mathfrak{W}_\sigma$ is the normal sub-semi-field $\mathfrak{J}(\mathfrak{G}_{0\sigma})$ and has a basis $w_\sigma^{G^*} = \sum_g (u^{n_1} + \cdots + u^{n_i})^g v^{(c_g^*)^\sigma G^*}$ for all G^* of $\mathfrak{G}^*$.

For every $\mathfrak{S}$ and $\mathfrak{T}$ we shall call the set $\mathfrak{W}_{-1}$ defined by $\mathfrak{S} \times \mathfrak{T} = \mathfrak{S} \times \mathfrak{W}_{-1}$ the product of $\mathfrak{S}$ and $\mathfrak{T}$. Then we prove

Lemma 1. *The product of $\mathfrak{S}$ and $\mathfrak{W}_{-1}$ is $\mathfrak{W}_{-2}$.*

For $\mathfrak{G}_{0,-1}$, the group of $\mathfrak{S}$, has a normal divisor $\mathfrak{N}$, such that $\mathfrak{G}_{0,-1}/\mathfrak{N} \cong \mathfrak{H}_c^*$, $\mathfrak{G}_{0,-1} = g_1(c_1^*)^{-1}\mathfrak{N} + g_2(c_2^*)^{-1}\mathfrak{N} + \cdots + g_j(c_j^*)^{-1}\mathfrak{N}$ and the resulting $\mathfrak{G}'_{0,-1} = (ng(c_g^*)^{-1}(c_g^*)^{-1}$, for all n and g) is equal to $\mathfrak{G}_{0,-2}$ since $ng(c_g^*)^{-1}(c_g^*)^{-1} = ng(c_g^*)^{-2}$. We shall make the induction part of this result later.

5. **Normal systems.** We now make the

Definition. *Let $\mathfrak{S}$ and $\mathfrak{S}'$ be normal semi-fields of order n with respective groups $\mathfrak{G} = (G)$ and $\mathfrak{G}' = (G')$. Then the pairs $\mathfrak{S}$, $\mathfrak{G}$ and $\mathfrak{S}'$, $\mathfrak{G}'$ are called equivalent, if $\mathfrak{G} \cong \mathfrak{G}'$ and if there exists a simple isomorphism*

$$s \leftrightarrow s' \qquad (s \text{ in } \mathfrak{S},\ s' \text{ in } \mathfrak{S}')$$

of $\mathfrak{S}$ and $\mathfrak{S}'$ such that

$$(s^G)' = s'^{G'}$$

for every s of $\mathfrak{S}$.

If the respective pairs $\mathfrak{S}$, $\mathfrak{G}$ and $\mathfrak{S}'$, $\mathfrak{G}'$; $\mathfrak{T}$, $\mathfrak{G}^*$ and $\mathfrak{T}'$, $\mathfrak{G}^{*\prime}$ are equivalent, the pairs $\mathfrak{S} \times \mathfrak{T}$, $\mathfrak{G} \times \mathfrak{G}^*$ and $\mathfrak{S}' \times \mathfrak{T}'$, $\mathfrak{G}' \times \mathfrak{G}^{*\prime}$ are equivalent.

The class of all equivalent pairs $\mathfrak{S}$, $\mathfrak{G}$ shall be designated by $(\mathfrak{S}, \mathfrak{G})$ and be called a **normal system of degree n with group $\mathfrak{G}$**. Then it follows that if $(\mathfrak{S}, \mathfrak{G}) = (\mathfrak{S}', \mathfrak{G}')$ and $(\mathfrak{T}, \mathfrak{G}^*) = (\mathfrak{T}', \mathfrak{G}^{*\prime})$, the corresponding systems $(\mathfrak{W}_\sigma, \mathfrak{G}^*)$ and $(\mathfrak{W}'_\sigma, \mathfrak{G}^{*\prime})$ of the previous section are equal, if we consider in forming $\mathfrak{W}'_\sigma$ the normal divisor $\mathfrak{N}'$ of $\mathfrak{G}'$ which corresponds to $\mathfrak{N}$ of $\mathfrak{G}$ in the isomorphism $\mathfrak{G} \cong \mathfrak{G}'$.

Theorem 7. *The set Σ of all normal systems with the same group $\mathfrak{G}$, having a normal divisor $\mathfrak{N}$ such that*

$$\mathfrak{G}/\mathfrak{N} \cong \mathfrak{H}_c,$$

$\mathfrak{H}_c$ a subgroup of the centrum of $\mathfrak{G}$, is closed with respect to the operation

$$(\mathfrak{S}, \mathfrak{G})(\mathfrak{T}, \mathfrak{G}^*) = (\mathfrak{W}_{-1}, \mathfrak{G}^*) \tag{3}$$

and

$$(\mathfrak{W}_{-\sigma}, \mathfrak{G}^*) = (\mathfrak{S}, \mathfrak{G})^\sigma (\mathfrak{T}, \mathfrak{G}^*). \tag{4}$$

The system $(\mathfrak{C}, \mathfrak{G})$, $\mathfrak{C}$ a diagonal algebra, is a left unit in Σ and every element in Σ has finite order.

For, using our hypothesis of $\mathfrak{G}$, we can form the product in the given sense, and this product is unique if we take a fixed normal divisor $\mathfrak{N}$ of $\mathfrak{G}$. Note that while the product $(\mathfrak{W}_{-1}, \mathfrak{G}^*)$ is unique there can exist other systems $(\mathfrak{S}', \mathfrak{G})$ such that $(\mathfrak{S}', \mathfrak{G})(\mathfrak{T}, \mathfrak{G}^*)=(\mathfrak{W}_{-1}, \mathfrak{G}^*)$. From the structure of $\mathfrak{W}_{-1}$ it is obvious that the product is not necessarily commutative.

We have $(\mathfrak{W}_{-1}, \mathfrak{G}^*)=(\mathfrak{S}, \mathfrak{G})(\mathfrak{T}, \mathfrak{G}^*)$. Assume (4) to be correct for all $\sigma \leqq \tau-1$, therefore $(\mathfrak{W}_{-(\tau-1)}, \mathfrak{G}^*)=(\mathfrak{S}, \mathfrak{G})^{\tau-1}(\mathfrak{T}, \mathfrak{G}^*)$. Then $(\mathfrak{S}, \mathfrak{G})=(\mathfrak{S}, \mathfrak{G}')$ with $\mathfrak{G}'=(n_g c_g^{*-(\tau-1)})$, and the product of $\mathfrak{S}$ and $\mathfrak{W}_{-(\tau-1)}$ is $\mathfrak{W}_{-\tau}$ from Lemma 1, $(\mathfrak{S}, \mathfrak{G})(\mathfrak{W}_{-(\tau-1)}, \mathfrak{G}^*)=(\mathfrak{W}_{-\tau}, \mathfrak{G}^*)=(\mathfrak{S}, \mathfrak{G})^{\tau}(\mathfrak{T}, \mathfrak{G}^*)$.

Let $(\mathfrak{E}, \mathfrak{G})(\mathfrak{S}, \mathfrak{G}^*)=(\mathfrak{W}_{-1}, \mathfrak{G}^*)$. Then if $\mathfrak{E}=(e^G)$, $\mathfrak{S}=(u^{G^*})$, $\mathfrak{W}_{-1}=(w^{G^*})$,

$$w^{G^*} = \sum_g (e^{n_1} + \cdots + e^{n_j})^g u^{(c_g^*)^{-1}G^*}.$$

Let $e^{n_1}+\cdots+e^{n_j}=f$; then f^g is a set of j orthogonal idempotents. But $\mathfrak{S}, \mathfrak{G}^*$ and $\mathfrak{W}_{-1}, \mathfrak{G}^*$ are equivalent. For $u^{G^*}\leftrightarrow w^{G^*}$ gives us this equivalence. Let $u^{G^*}u^{H^*}=\sum_{F^*}\rho_{F^*}^{(G,H)}u^{F^*}$ ($\rho_{F^*}^{(G,H)}$ in $\mathfrak{K}$); then

$$\begin{aligned}
w^{G^*}w^{H^*} &= \sum_g (f^g)u^{G^*(c_g^*)^{-1}} \sum_h (f^h)u^{H^*(c_h^*)^{-1}} \\
&= \sum_g \sum_h (f^g f^h)u^{G^*(c_g^*)^{-1}}u^{H^*(c_h^*)^{-1}} \\
&= \sum_g f^g(u^{G^*}u^{H^*})^{(c_g^*)^{-1}} \\
&= \sum_g f^g \sum_{F^*} \rho_{F^*}^{(G,H)}u^{F^*(c_g^*)^{-1}} \\
&= \sum_{F^*} \rho_{F^*}^{(G,H)} \sum_g (f^g)u^{F^*(c_g^*)^{-1}} \\
&= \sum_{F^*} \rho_{F^*}^{(G,H)} w^{F^*}.
\end{aligned}$$

Hence $(\mathfrak{S}, \mathfrak{G}^*)=(\mathfrak{W}_{-1}, \mathfrak{G}^*)$ and $(\mathfrak{E}, \mathfrak{G})(\mathfrak{S}, \mathfrak{G}^*)=(\mathfrak{S}, \mathfrak{G}^*)$.

From (4) we have

$$(\mathfrak{W}_{-j}, \mathfrak{G}^*) = (\mathfrak{S}, \mathfrak{G})^j(\mathfrak{T}, \mathfrak{G}^*);$$

$\mathfrak{W}_{-j}$ consists of all elements of $\mathfrak{S}\times\mathfrak{T}$ unaltered by $ng(c_g^*)^{-j}$. But j is the order of $\mathfrak{H}_c$ and $(c_g^*)^{-j}$ is the unity of $\mathfrak{G}^*$, $\mathfrak{W}_{-j}=\mathfrak{T}$, $(\mathfrak{T}, \mathfrak{G}^*)=(\mathfrak{S}, \mathfrak{G})^j(\mathfrak{T}, \mathfrak{G}^*)$. Hence every element has finite order in Σ. It does not follow, however, that $(\mathfrak{S}, \mathfrak{G})^j$ is equal to $(\mathfrak{E}, \mathfrak{G})$. For let $\mathfrak{Y}$ be a normal semi-field such that $\mathfrak{Y}=\mathfrak{Y}(\mathfrak{H}_c)\times\mathfrak{F}$ with $\mathfrak{F}$ a diagonal algebra and $\mathfrak{Y}(\mathfrak{H}_c)$ an arbitrary normal semi-field with group $\mathfrak{N}$; then a simple computation shows that $(\mathfrak{Y}, \mathfrak{G})(\mathfrak{T}, \mathfrak{G}^*)=(\mathfrak{T}, \mathfrak{G}^*)$ for all $(\mathfrak{T}, \mathfrak{G}^*)$ of Σ.

THEOREM 8. *The associative law holds in the set Σ of Theorem 7 if and only if*

$$\mathfrak{G} = \mathfrak{N} \times \mathfrak{H}_c.$$

For let

$$(5)\qquad [(\mathfrak{S}, \mathfrak{G})(\mathfrak{T}, \mathfrak{G}^*)](\mathfrak{R}, \mathfrak{G}^{**}) = (\mathfrak{W}, \mathfrak{G}^*)(\mathfrak{R}, \mathfrak{G}^{**}) = (\mathfrak{W}_0, \mathfrak{G}^{**}).$$

Then $\mathfrak{W}_0$ is the normal sub-semi-field of $\mathfrak{S}\times\mathfrak{T}\times\mathfrak{R}$ unaltered by the group $(ng(c_\sigma^*)^{-1})\times(n^*g^*(c_\rho^{**})^{-1})$. On the other hand

$$(6)\qquad (\mathfrak{S}, \mathfrak{G})[(\mathfrak{T}, \mathfrak{G}^*)(\mathfrak{R}, \mathfrak{G}^{**})] = (\mathfrak{S}, \mathfrak{G})(\mathfrak{W}_1, \mathfrak{G}^{**}) = (\mathfrak{W}_{10}, \mathfrak{G}^{**}).$$

$\mathfrak{W}_{10}$ is the normal sub-semi-field of $\mathfrak{S}\times\mathfrak{T}\times\mathfrak{R}$ unaltered by the group $(n^*g^*(c_\rho^{**})^{-1})\times(ng(c_\rho^{**})^{-1})$.

The associative law holds if $(\mathfrak{W}_0, \mathfrak{G}^{**})=(\mathfrak{W}_{10}, \mathfrak{G}^{**})$, in other words if $\mathfrak{W}_0, \mathfrak{G}^{**}$ is equivalent to $\mathfrak{W}_{10}, \mathfrak{G}^{**}$. For this $\mathfrak{W}_0$ must be isomorphic to $\mathfrak{W}_{10}$. Let us consider the special case where $\mathfrak{S}$, $\mathfrak{T}$, $\mathfrak{R}$ are normal fields whose intersections are $\mathfrak{K}$. Then $\mathfrak{W}_0$ and $\mathfrak{W}_{10}$ are normal fields since they are sub-fields of the field $\mathfrak{S}\times\mathfrak{T}\times\mathfrak{R}$. Then we have $\mathfrak{W}_0$ isomorphic to $\mathfrak{W}_{10}$ if and only if the subgroups of $\mathfrak{G}\times\mathfrak{G}^*\times\mathfrak{G}^{**}$ whose elements leave their elements unaltered are conjugate subgroups of $\mathfrak{G}\times\mathfrak{G}^*\times\mathfrak{G}^{**}$(10). But these subgroups are normal divisors of $\mathfrak{G}\times\mathfrak{G}^*\times\mathfrak{G}^{**}$, hence $\mathfrak{W}_0\cong\mathfrak{W}_{10}$ if and only if their corresponding groups are equal and $\mathfrak{W}_0=\mathfrak{W}_{10}$. The equality of the groups $(ng(c_\sigma^*)^{-1})\times(n^*g^*(c_\rho^{**})^{-1})$ and $(ng(c_\sigma^{**})^{-1})\times(n^*g^*(c_\rho^{**})^{-1})$ includes, of course, the equivalence of $\mathfrak{W}_0$, $\mathfrak{G}^{**}$ and $\mathfrak{W}_{10}$, $\mathfrak{G}^{**}$ in general. Thus the associative law holds in Σ if and only if

$$(7)\qquad (ng(c_\sigma^*)^{-1}) \times (n^*g^*(c_{\sigma^*}^{**})^{-1}) = (ng(c_\sigma^{**})^{-1}) \times (n^*g^*(c_{\rho^*}^{**})^{-1}).$$

If (7) holds we have $g(c_\sigma^*)^{-1}$ in the left member of (7) and

$$g(c_\sigma^*)^{-1} = ng(c_\sigma^{**})^{-1}n^*h^*(c_{h^*}^{**})^{-1}.$$

Hence n and n^* are the unities of $\mathfrak{N}$ and $\mathfrak{N}^*$ respectively, $(c_{h^*}^{**})^{-1}=c_\sigma^{**}$, $h^*=g^{*-1}$, and $g(c_\sigma^*)^{-1}=gg^{*-1}$ and consequently

$$(8)\qquad c_\sigma^* = g^*.$$

But this means $\mathfrak{G}=c_1\mathfrak{N}+c_2\mathfrak{N}+\cdots+c_j\mathfrak{N}$. Since the c_ν are in the centrum of $\mathfrak{G}$ and $ij=n$, i, j, n the orders of $\mathfrak{N}$, $\mathfrak{H}_c$, $\mathfrak{G}$ respectively, it follows(11) that $\mathfrak{G}=\mathfrak{N}\times\mathfrak{H}_c$.

Our necessary condition (8) is sufficient. For if (8) holds, $\mathfrak{W}_0$ in (5) is the subset of $\mathfrak{S}\times\mathfrak{T}\times\mathfrak{R}$ unaltered by $(ncc^{*-1})\times(n^*c^{*\prime}c^{**\prime-1})$ and $\mathfrak{W}_{10}$ the subset unaltered by $(ncc^{**-1})\times(n^*c^{*\prime}c^{**\prime-1})$. But $ncc^{*-1}n^*c^{*\prime}c^{**\prime-1}$ $=ncc^{**-1}n^*(c^{*-1}c^{*\prime})[c^{**}(c^{**\prime})^{-1}]$; hence the groups are equal and therefore $(\mathfrak{W}_0, \mathfrak{G}^{**})=(\mathfrak{W}_{10}, \mathfrak{G}^{**})$. This proves the theorem.

As an immediate consequence we have the

Corollary. *Let* $\mathfrak{G}=\mathfrak{N}\times\mathfrak{H}_c$; *then* $(\mathfrak{S}, \mathfrak{G})^j$ *is a left and right unit of* $(\mathfrak{S}, \mathfrak{G})$.

(10) Albert, *Modern Higher Algebra*, p. 176, Theorem 6.

(11) Ibid., p. 127.

For the associative law holds and

$$(\mathfrak{S}, \mathfrak{G}) = (\mathfrak{S}, \mathfrak{G})^i(\mathfrak{S}, \mathfrak{G}) = (\mathfrak{S}, \mathfrak{G})(\mathfrak{S}, \mathfrak{G})^{i-1}(\mathfrak{S}, \mathfrak{G}) = (\mathfrak{S}, \mathfrak{G})(\mathfrak{S}, \mathfrak{G})^i.$$

In general there does not exist a right unit for all elements in Σ. For if $(\mathfrak{B}, \mathfrak{G}^*)$ were a right unit, $(\mathfrak{S}, \mathfrak{G})(\mathfrak{B}, \mathfrak{G}^*) = (\mathfrak{S}, \mathfrak{G}^*)$ for all $(\mathfrak{S}, \mathfrak{G})$, we would have in particular $(\mathfrak{E}, \mathfrak{G})(\mathfrak{B}, \mathfrak{G}^*) = (\mathfrak{E}, \mathfrak{G}^*)$; but $(\mathfrak{E}, \mathfrak{G})(\mathfrak{B}, \mathfrak{G}^*) = (\mathfrak{B}, \mathfrak{G}^*)$; hence $(\mathfrak{E}, \mathfrak{G}^*)$ would be the right unit. Now let $\mathfrak{S}$ be a field. Then $(\mathfrak{S}, \mathfrak{G})(\mathfrak{E}, \mathfrak{G}^*) = (\mathfrak{W}_{-1}, \mathfrak{G}^*)$, and if $\mathfrak{S} = (u^G)$, $\mathfrak{E} = (e^{G^*})$, $\mathfrak{W}_{-1} = (w^{G^*})$, then

$$w^{n^*c^{*\prime}} = \sum_c (u^{n_1} + \cdots + u^{n_i})^c e^{c^{*-1}n^*c^{*\prime}}$$

and

$$\begin{aligned} ww^{n^*} &= \sum_c (u^{n_1} + \cdots + u^{n_i})^c e^{c^{*-1}} \sum_{c'} (u^{n_1} + \cdots + u^{n_i})^{c'} e^{c^{*\prime-1}n^*} \\ &= \sum_c \sum_{c'} (u^{n_1} + \cdots + u^{n_i})^c (u^{n_1} + \cdots + u^{n_i})^{c'} e^{c^{*-1}} e^{c^{*\prime-1}n^*} \\ &= 0 \end{aligned}$$

since every $e^{c^{*-1}}e^{c^{*\prime-1}n^*} = 0$. Thus $\mathfrak{W}_{-1}$ contains divisors of zero and cannot be equivalent to the field $\mathfrak{S}$; $(\mathfrak{E}, \mathfrak{G})$ is not a right unit—a contradiction.

Let now $\mathfrak{G}$ be abelian. Then $\mathfrak{G}$ is its own centrum and we can consider the factorization $\mathfrak{S} \times \mathfrak{T} = \mathfrak{S} \times \mathfrak{W}_{-1}$ deriving from the isomorphism $\mathfrak{G} \cong \mathfrak{G}^*$. We do this and have

THEOREM 9. *The set Σ of all normal systems of order n with the same abelian group $\mathfrak{G}$ forms an abelian group with respect to the operation*

$$(\mathfrak{S}, \mathfrak{G})(\mathfrak{T}, \mathfrak{G}^*) = (\mathfrak{W}_{-1}, \mathfrak{G}^*)$$

and

$$(\mathfrak{W}_{-r}, \mathfrak{G}^*) = (\mathfrak{S}, \mathfrak{G})^r(\mathfrak{T}, \mathfrak{G}^*).$$

The identity element in Σ is the system $(\mathfrak{E}, \mathfrak{G})$ where $\mathfrak{E}$ is diagonal and every element of Σ has finite order.

We only have to show that the multiplication in Σ is commutative, for all the other properties are then immediate consequences of Theorems 7 and 8. Let $\mathfrak{S} = (u^G)$, $\mathfrak{T} = (v^{G^*})$. Then $\mathfrak{W}_{-1} = (w^{G^*})$ with $w = \sum_G u^G v^{G^{*-1}}$ and from this it follows that $(\mathfrak{S}, \mathfrak{G})(\mathfrak{T}, \mathfrak{G}^*) = (\mathfrak{T}, \mathfrak{G}^*)(\mathfrak{S}, \mathfrak{G})$.

This theorem can also be derived as a consequence of the cyclic case.

Let now $\mathfrak{G}$ be cyclic of order n. $\mathfrak{G} = [S]$ has the identity as normal divisor and $\mathfrak{G} = [I] \times [S]$ and we can form products of two normal semi-fields with $\mathfrak{G}$ as group. Albert discusses this case[12]. But in case the order n of $\mathfrak{G}$ is not a prime, $n = ij$, we have a subgroup $[S^i]$ of $\mathfrak{G}$, which is, of course, a normal divi-

[12] *Structure of Algebras*, 1939.

sor of order j of $\mathfrak{G}$. Now $\mathfrak{G}=[S^i]+S[S^i]+\cdots+S^{i-1}[S^i]$ and $\mathfrak{G}/[S^i]\cong[S^i]$. Consequently we can consider groups $\mathfrak{G}_{0,-\sigma}$, $\sigma=1,\cdots,i-1$, arising from this relation. $\mathfrak{G}_{0,-\sigma}$ consists of elements of the form $S^{i\nu}S^{\mu}S^{*-i\sigma\mu}$. It is easily seen that $\mathfrak{G}_{0,-\sigma}$ is a cyclic group of order n generated by $SS^{*-i\sigma}$. Hence $\mathfrak{G}_{0,-\sigma}$ is the same as the group $\mathfrak{G}'_{0,-j\sigma}$ arising from the isomorphism $\mathfrak{G}=\mathfrak{G}/[I]\cong\mathfrak{G}^*$, and we do not get any new direct factorizations $\mathfrak{S}\times\mathfrak{W}$ by considering $\mathfrak{G}_{0,-\sigma}$.

Note that we considered only factorizations $\mathfrak{S}\times\mathfrak{T}=\mathfrak{S}\times\mathfrak{W}$ which originated by factorizations of the group $\mathfrak{G}\times\mathfrak{G}^*=\mathfrak{G}_0\times\mathfrak{G}^*$. But there can exist different factorizations of $\mathfrak{S}\times\mathfrak{T}$. Since we are dealing with a generalization of the cyclic case, that is, direct factorizations of $\mathfrak{S}\times\mathfrak{T}$ associated with direct factorizations of the group $\mathfrak{G}\times\mathfrak{G}^*$, there remains the problem of finding all other possible factorizations.

6. **Generalized crossed products.** Teichmüller defined crossed products of normal semi-fields with their groups similarly to the definition of crossed products of normal fields([13]).

Let $\mathfrak{S}$ be a normal semi-field of order n and $\mathfrak{G}$ an automorphism group of $\mathfrak{S}$. To every two elements G, H of $\mathfrak{G}$ there shall correspond an element $a_{G,H}$ of $\mathfrak{S}$ such that

$$a_{F,H}a_{G,FH} = a_{G,F}^{H}a_{GF,H}. \tag{9}$$

Furthermore, to every G of $\mathfrak{G}$ there corresponds a symbol x_G satisfying

$$sx_G = x_G s^G \qquad (s \text{ in } \mathfrak{S}), \tag{10}$$

$$x_G x_H = x_{GH}a_{G,H}. \tag{11}$$

Then the set

$$A = (\mathfrak{S}, \mathfrak{G}, a) = \sum_G \mathfrak{S}x_G$$

is a normal simple algebra of order n^2 over $\mathfrak{K}$([14]), and $A=(\mathfrak{S}, \mathfrak{G}, a)$ is called the crossed product of $\mathfrak{S}$ with its group $\mathfrak{G}$ and factor set a.

Teichmüller proved that if $\mathfrak{S}=\mathfrak{Q}\times\mathfrak{C}$,

$$(\mathfrak{S}, \mathfrak{G}, a) \sim (\mathfrak{Q}, \mathfrak{H}, a')$$

with $\mathfrak{H}$ the subgroup of $\mathfrak{G}$ belonging to $\mathfrak{Q}$ and a' the subset of a corresponding to pairs of elements of $\mathfrak{H}$. Since the crossed product of a normal field with its group is a total matric algebra if the factor set $a=i$, that is, all $a_{G,H}=1$, we have

$$(\mathfrak{S}, \mathfrak{G}, i) \sim (\mathfrak{Q}, \mathfrak{H}, i) \sim 1.$$

It is easily seen that all theorems which hold for factor sets of crossed

([13]) Teichmüller, p. 93.
([14]) Ibid., p. 96.

products of normal fields[15] hold for those of normal semi-fields. In particular those about associated factor sets and that $(\mathfrak{S}, \mathfrak{G}, a) \times (\mathfrak{S}, \mathfrak{G}, b) \sim (\mathfrak{S}, \mathfrak{G}, ab)$ with the equivalence of algebras, that is, $\mathfrak{A} \sim \mathfrak{B}$ if $\mathfrak{A} = \mathfrak{B} \times \mathfrak{M}$ where $\mathfrak{M}$ is a total matric algebra, and ab is the factor set $c_{GE^*, HF^*} = a_{G,H} b_{E^*, F^*}$ hold.

From a theorem by Teichmüller[16] we know that, for two crossed products $A = (\mathfrak{S}, \mathfrak{G}, a)$ and $B = (\mathfrak{T}, \mathfrak{G}^*, b)$ of normal semi-fields $\mathfrak{S}$ and $\mathfrak{T}$ with their respective groups $\mathfrak{G}$ and $\mathfrak{G}^*$, the direct product $J = A \times B$ is a crossed product of the normal semi-field $\mathfrak{S} \times \mathfrak{T}$ with its group $\mathfrak{G} \times \mathfrak{G}^*$ and the product ab of the factor sets a and b a factor set:

$$J = A \times B = (\mathfrak{S}, \mathfrak{G}, a) \times (\mathfrak{T}, \mathfrak{G}^*, b) = (\mathfrak{S} \times \mathfrak{T}, \mathfrak{G} \times \mathfrak{G}^*, ab).$$

Note that what we have called the factor set $ab = d$ in $\mathfrak{S} \times \mathfrak{T}$ is that factor set consisting of $d_{K,L} = a_{G,H} b_{E^*, F^*}$ for every $K = GE^*$, $L = HF^*$ in $\mathfrak{G} \times \mathfrak{G}^*$, with G, H in $\mathfrak{G}$ and E^*, F^* in $\mathfrak{G}^*$.

Let $\mathfrak{G}$ and $\mathfrak{G}^*$ now be as in Theorem 6. Then we know from Lemma 1 that $\mathfrak{W}_{-1}$ is contained in J. Let us consider the possibility of a factorization of J as direct product of crossed products of $\mathfrak{S}$ and $\mathfrak{W}_{-1}$ with their respective groups $\mathfrak{G}_0$ and $\mathfrak{G}^*$ with b as factor set for $\mathfrak{W}_{-1}$, that is, a factorization

$$J = A_1 \times B_1 = (\mathfrak{S}, \mathfrak{G}_0, c) \times (\mathfrak{W}_{-1}, \mathfrak{G}^*, b).$$

From Teichmüller's theorem we know that

$$\begin{aligned} (\mathfrak{S}, \mathfrak{G}_0, c) \times (\mathfrak{W}_{-1}, \mathfrak{G}^*, b) &= (\mathfrak{S} \times \mathfrak{W}_{-1}, \mathfrak{G}_0 \times \mathfrak{G}^*, cb) \\ &= (\mathfrak{S} \times \mathfrak{T}, \mathfrak{G} \times \mathfrak{G}^*, ab) \\ &= (\mathfrak{S}, \mathfrak{G}, a) \times (\mathfrak{T}, \mathfrak{G}^*, b). \end{aligned} \tag{12}$$

THEOREM 10. *Let* $J = (\mathfrak{S}, \mathfrak{G}, a) \times (\mathfrak{T}, \mathfrak{G}^*, b)$ *and* $\mathfrak{G}$, $\mathfrak{G}^*$ *be as in Theorem* 6. *Then* J *has the factorization*

$$J = (\mathfrak{S}, \mathfrak{G}_0, c) \times (\mathfrak{W}_{-1}, \mathfrak{G}^*, b) \tag{13}$$

if and only if

(a) *the factor set* b *is in* $\mathfrak{T}(\mathfrak{H}_c^*) = \mathfrak{T} \cap \mathfrak{W}_{-1}$,

(b) *there exist* n *elements* r_{G_0} *in* $\mathfrak{S} \times \mathfrak{T}$, *such that*

$$b_{(c_g^*)^{-1}, (c_h^*)^{-1}} R_{G_0, H_0} \quad in \ \mathfrak{S}, \tag{14}$$

$$b_{(c_g^*)^{-1}, G^*} r_{G_0}^{G^*} = b_{G^*, (c_g^*)^{-1}} r_{G_0}, \tag{15}$$

with

$$c_{G_0, H_0} = a_{ng, n'h} b_{(c_g^*)^{-1}, (c_h^*)^{-1}} R_{G_0 H_0}, \tag{16}$$

where

---

[15] Albert, A. A., *Structure of Algebras*, chap. 5.

[16] Teichmüller, p. 100.

$$R_{G_0,H_0} = \frac{r_{G_0}^{H_0} r_{H_0}}{r_{G_0 H_0}}. \tag{17}$$

That the factor set b in $\mathfrak{T}$ must consist of elements in $\mathfrak{T} \cap \mathfrak{W}_{-1}$ is trivial since its elements must lie in $\mathfrak{W}_{-1}$ in order that the crossed product $(\mathfrak{W}_{-1}, \mathfrak{G}^*, b)$ has meaning. But $\mathfrak{W}_{-1}$ consists of all elements of $\mathfrak{S}\times\mathfrak{T}$ unaltered by $\mathfrak{G}_0$. The automorphisms of $\mathfrak{G}_0$ are products of automorphisms of $\mathfrak{G}$ with automorphisms of $\mathfrak{H}_c^*$. Since the elements of $\mathfrak{T}$ are unaltered by $\mathfrak{G}$, $\mathfrak{T} \cap \mathfrak{W}_{-1}$ consists of all elements of $\mathfrak{T}$ unaltered by $\mathfrak{H}_c^*$. $\mathfrak{H}_c^*$ is a normal divisor of $\mathfrak{G}^*$, and from Theorem 1 we have $\mathfrak{W}_{-1} \cap \mathfrak{T} = \mathfrak{T}(\mathfrak{H}_c^*)$.

To prove that the conditions for the factor set c are necessary let us assume that there exists a factor set c such that (13) holds. Let $(\mathfrak{S}, \mathfrak{G}_0, c) = (u^{G_0} z_{H_0})$, $(\mathfrak{S}, \mathfrak{G}, a) = (u^G x_H)$, $(\mathfrak{T}, \mathfrak{G}^*, b) = (v^{G^*} y_{H^*})$ where the u, z; u, x; v, y; a, b, c satisfy formulas similar to (9), (10) and (11). Then we have the following equations in J

$$u z_{G_0} = z_{G_0} u^G, \qquad w z_{G_0} = z_{G_0} w. \tag{18}$$

Now an arbitrary element of $\mathfrak{G}_0$ has the form $ng(c_g^*)^{-1}$ and we write $ng(c_g^*)^{-1} = G_0$; then we have in J

$$u(x_{ng} y_{(c_g^*)^{-1}})^{-1} z_{G_0} = (x_{ng} y_{(c_g^*)^{-1}})^{-1} z_{G_0} u, \qquad w(x_{ng} y_{(c_g^*)^{-1}})^{-1} z_{G_0} = (x_{ng} y_{(c_g^*)^{-1}})^{-1} z_{G_0} w,$$

and hence since v is in $\mathfrak{S}\times\mathfrak{W}_{-1} = \mathfrak{S}\times\mathfrak{T}$,

$$v(x_{ng} y_{(c_g^*)^{-1}})^{-1} z_{G_0} = (x_{ng} y_{(c_g^*)^{-1}})^{-1} z_{G_0} v.$$

Consequently

$$(x_{ng} y_{(c_g^*)^{-1}})^{-1} z_{G_0} = r_{G_0} \text{ in } \mathfrak{S} \times \mathfrak{T}, \tag{19}$$

since the only elements of J commutative with all elements of $\mathfrak{S}\times\mathfrak{T}$ are in $\mathfrak{S}\times\mathfrak{T}$. From (19) it follows that

$$z_{G_0} = (x_{ng} y_{(c_g^*)^{-1}}) r_{G_0}.$$

Now

$$\begin{aligned} z_{G_0} z_{H_0} &= z_{G_0 H_0} c_{G_0,H_0} \\ &= x_{ng} y_{(c_g^*)^{-1}} r_{G_0} x_{n'h} y_{(c_h^*)^{-1}} r_{H_0} \\ &= x_{ngn'h} y_{(c_g^*)^{-1}(c_h^*)^{-1}} a_{ng,n'h} b_{(c_g^*)^{-1},(c_h^*)^{-1}} r_{G_0}^{H_0} r_{H_0} \\ &= x_{ngn'h} y_{(c_g^*)^{-1}(c_h^*)^{-1}} r_{G_0 H_0} c_{G_0,H_0}. \end{aligned}$$

Hence we get (16)

$$\begin{aligned} c_{G_0,H_0} &= a_{ng,n'h} b_{(c_g^*)^{-1},(c_h^*)^{-1}} \frac{r_{G_0}^{H_0} r_{H_0}}{r_{G_0 H_0}} \\ &= a_{ng,n'h} b_{(c_g^*)^{-1},(c_h^*)^{-1}} R_{G_0,H_0}. \end{aligned}$$

The fraction R_{G_0,H_0} designates merely the result of a transformation of the basis in J. The c's form a factor set since we can derive them by a transformation of a basis of J. But the sets c and a are in $\mathfrak{S}$ and we have (14)

$$b_{(c_g^*)^{-1},(c_h^*)^{-1}} R_{G_0,H_0} \text{ in } \mathfrak{S} \qquad \text{for all } G_0, H_0.$$

In addition, since we want to form the direct product $(\mathfrak{S}, \mathfrak{G}_0, c)\times(\mathfrak{W}_{-1}, \mathfrak{G}^*, b)$ we must have

$$z_{G_0} y_{G^*} = y_{G^*} z_{G_0} = x_{n_g} y_{(c_g^*)^{-1}} r_{G_0} y_{G^*} = y_{G^*} x_{n_g} y_{(c_g^*)^{-1}} r_{G_0}. \tag{20}$$

The x and y are commutative and consequently we have (15)

$$b_{(c_g^*)^{-1},G^*} r_{G_0}^{G^*} = b_{G^*,(c_g^*)^{-1}} r_{G_0}.$$

The conditions are sufficient. For with c as in (16) we have $(\mathfrak{S}, \mathfrak{G}_0, c)\times(\mathfrak{W}_{-1}, \mathfrak{G}^*, b)=(\mathfrak{S}\times\mathfrak{W}_{-1}, \mathfrak{G}_0\times\mathfrak{G}^*, cb)$ and cb is a factor set in $S\times\mathfrak{W}_{-1}=\mathfrak{S}\times\mathfrak{T}$ associated with ab and therefore $(\mathfrak{S}, \mathfrak{G}_0, c)\times(\mathfrak{W}_{-1}, \mathfrak{G}^*, b)=(S\times\mathfrak{W}_{-1}, \mathfrak{G}_0\times\mathfrak{G}^*, cb)=(\mathfrak{S}\times\mathfrak{T}, \mathfrak{G}\times\mathfrak{G}^*, ab)=(\mathfrak{S}, \mathfrak{G}, a)\times(\mathfrak{T}, \mathfrak{G}^*, b)$. This proves Theorem 10.

Note that the factor set c consists of elements which are products of elements of a with elements of that part of b which corresponds to pairs of automorphisms of $\mathfrak{H}_c^*$. By a change of the basis of J these products are multiplied by R_{G_0,H_0} such that c is in $\mathfrak{S}$.

In case $\mathfrak{G}^*=\mathfrak{N}^{*\prime}\times\mathfrak{H}_c^*$ we have $\mathfrak{T}=\mathfrak{T}(\mathfrak{N}^{*\prime})\times\mathfrak{T}(\mathfrak{H}_c^*)$. Then the factor set b may be such that $(\mathfrak{T}, \mathfrak{G}^*, b)=(\mathfrak{T}(\mathfrak{N}^{*\prime}), \mathfrak{H}_c^*, b')\times(\mathfrak{T}(\mathfrak{H}_c^*), \mathfrak{N}^{*\prime}, b'')$. But if the factor set b satisfies the first condition, the b' are in $\mathfrak{K}$, since they are in $\mathfrak{T}(\mathfrak{N}^{*\prime})$ and $\mathfrak{T}(\mathfrak{H}_c^*)$ and their intersection is $\mathfrak{K}$. This gives

THEOREM 11. *If $\mathfrak{G}=\mathfrak{N}^{*\prime}\times\mathfrak{H}_c^*$, $\mathfrak{G}/\mathfrak{N}\cong\mathfrak{H}_c^*$ is as in Theorem 6, the factor set b is in $\mathfrak{T}(\mathfrak{H}_c^*)$, and $(\mathfrak{T}, \mathfrak{G}^*, b)=(\mathfrak{T}(\mathfrak{N}^{*\prime}), \mathfrak{H}_c^*, b')\times(\mathfrak{T}(\mathfrak{H}_c^*), \mathfrak{N}^{*\prime}, b'')$, then*

$$J = (\mathfrak{S}, \mathfrak{G}, a) \times (\mathfrak{T}, \mathfrak{G}^*, b) = (\mathfrak{S}, \mathfrak{G}_0, c) \times (\mathfrak{W}_{-1}, \mathfrak{G}^*, b)$$

with

$$c_{G_0H_0} = a_{n_g,n'_h} b_{(c_g^*)^{-1},(c_h^*)^{-1}}.$$

We can consider the crossed product $(\mathfrak{S}, \mathfrak{G}, a)$ as defined by the normal system $\mathfrak{A}=(\mathfrak{S}, \mathfrak{G})$, if the factor set a is in the intersection of all $\mathfrak{S}$ in $\mathfrak{A}$. Now we know from Theorem 8 that if $\mathfrak{S}=(u^{G_\nu}, \nu=1, \cdots, n)$ with group $\mathfrak{G}$ having normal divisor $\mathfrak{N}$ such that $\mathfrak{G}/\mathfrak{N}\cong\mathfrak{H}_c$, a subgroup of the centrum of $\mathfrak{G}$, and $\mathfrak{S}^*=(v^{G_\nu^*})$, $v=\sum_{\sigma J} f^\sigma u^{(c_\sigma^*)^{-1}}$, f^σ a set of j orthogonal idempotents, then the pairs $\mathfrak{S}$, $\mathfrak{G}$ and $\mathfrak{S}^*$, $\mathfrak{G}^*$ are equivalent. An easy computation shows that the intersection $\mathfrak{S} \cap \mathfrak{S}^*$ is $\mathfrak{S}(\mathfrak{H}_c)$, the sub-semi-field of $\mathfrak{S}$ unaltered by $\mathfrak{H}_c$. Thus it is at least necessary that the factor set a is in $\mathfrak{S}(\mathfrak{H}_c)$.

It follows that if $\mathfrak{G}$ is abelian the intersection of all $\mathfrak{S}$ in $\mathfrak{A}$ must be $\mathfrak{S}(\mathfrak{G})$ and is therefore $\mathfrak{K}$.

In this fashion we write $(\mathfrak{A}, a)$ with $\mathfrak{A}$ a normal system $(\mathfrak{S}, \mathfrak{G})$ and a a factor set in the intersection of all $\mathfrak{S}$ in $\mathfrak{A}$. We say for brevity that a is in $\mathfrak{A}$. Then as a consequence of Theorem 10 we have

THEOREM 12. *Let $\mathfrak{A}$ and $\mathfrak{B}$ be normal systems with the same group $\mathfrak{G}$ having the property of Theorem 7. Suppose that a and b are respective factor sets in $\mathfrak{A}$ and $\mathfrak{B}$ such that*

(a) *the factor set b is in $\mathfrak{B}$ and $\mathfrak{A}\mathfrak{B}$,*

(b) *there exist n elements r_{G_0} in $\mathfrak{A}\mathfrak{B}$, such that*

$$b_{(c_g^*)^{-1},(c_h^*)^{-1}} R_{G_0,H_0} \text{ is in } \mathfrak{A},$$

and

$$b_{(c_g^*)^{-1},G^*} \overset{G^*}{r_{G_0}} = b_{G^*,(c_g^*)^{-1}} r_{G_0},$$

with the factor set c as in (16).

Then

$$(\mathfrak{A}, a) \times (\mathfrak{B}, b) = (\mathfrak{A}, c) \times (\mathfrak{A}\mathfrak{B}, b).$$

COROLLARY I. *If the factor set b is in the reference field, and*

$$b_{G^*,(c_g^*)^{-1}} = b_{(c_g^*)^{-1},G^*},$$

then $(\mathfrak{A}, a)\times(\mathfrak{B}, b)=(\mathfrak{A}, c)\times(\mathfrak{A}\mathfrak{B}, b)$ with $c=\{c_{G_0,H_0}\}$,

$$c_{G_0,H_0} = a_{n_g,n'_h} b_{(c_g^*)^{-1},(c_h^*)^{-1}}.$$

COROLLARY II. *Let*

$$a_{G,c_g} = a_{c_g,G}.$$

Then

$$(\mathfrak{A}, a)^2 = (\mathfrak{A}, c) \times (\mathfrak{A}^2, a)$$

with $c=\{c_{G_0,H_0}\}$.

We now discuss the case where $\mathfrak{G}$ is abelian. Then we saw that in order to form a crossed product of the system $\mathfrak{A}$ with its group $\mathfrak{G}$ the factor set must be in $\mathfrak{K}$. We then have

THEOREM 13. *Let $\mathfrak{A}$ and $\mathfrak{B}$ be abelian systems with group $\mathfrak{G}$, a and b two factor sets in the reference field, and*

$$b_{G^*,H^*} = b_{H^*,G^*}; \tag{21}$$

then $(\mathfrak{A}, a)\times(\mathfrak{B}, b)=(\mathfrak{A}, ab^{-1})\times(\mathfrak{A}\mathfrak{B}, b)$.

This follows from Theorem 12, for the three conditions are satisfied with $r_{G_0}=1$.

As a consequence we have

THEOREM 14. *If $\mathfrak{A}$ is an abelian system and*

$$a_{G,H} = a_{H,G}, \tag{22}$$

then

$$(\mathfrak{A}, a)^{\sigma} \sim (\mathfrak{A}^{\sigma}, a). \tag{23}$$

For from Theorem 14 we have $(\mathfrak{A}, a)^2 = (\mathfrak{A}, aa^{-1}) \times (\mathfrak{A}^2, a)$ with $a_{G,H}a_{G,H}^{-1} = 1$ and hence $(\mathfrak{A}, aa^{-1})$ is a total matric algebra; this is (23) for $\sigma = 2$. In general (23) follows then by repeated use of this argument.

If the group $\mathfrak{G}$ of $\mathfrak{A}$ is abelian and the direct product of not more than two cyclic groups we can show that the crossed product $(\mathfrak{A}, a)$ is the direct product of two cyclic algebras. It is not known whether a similar statement can be made for the case where the group of $\mathfrak{A}$ is the direct product of more than two cyclic factors unless (22) holds, and hence that hypothesis seems to be necessary in order that a result of the type given above may be obtained.

UNIVERSITY OF CHICAGO,
CHICAGO, ILL.

THE ACCURACY OF THE GAUSSIAN APPROXIMATION TO THE SUM OF INDEPENDENT VARIATES

BY
ANDREW C. BERRY

1. Introduction

The sum of finitely many variates possesses, under familiar conditions, an almost Gaussian probability distribution. This already much discussed "central limit theorem"(1) in the theory of probability is the object of further investigation in the present paper. The cases of Liapounoff(2), Lindeberg(3), and Feller(4) will be reviewed. Numerical estimates for the degrees of approximation attained in these cases will be presented in the three theorems of §4. Theorem 3, the arithmetical refinement of the general theorem of Feller, constitutes our principal result. As the foregoing implies, we require throughout the paper that the given variates be totally independent. And we consider only one-dimensional variates.

The first three sections of the paper are devoted to the preparatory Theorem 1 in which the variates meet the further condition of possessing finite third order absolute moments. Let $X_1, X_2, \cdots, X_n$ be the given variates. For each k $(k=1, 2, \cdots, n)$ let $\mu_2(X_k)$ and $\mu_3(X_k)$ denote, respectively, the second and third order absolute moments of X_k about its mean (expected) value α_k. These moments are either both zero or both positive. The former case arises only when X_k is essentially constant, i.e., differs from its mean value at most in cases of total probability zero. To avoid trivialities we suppose that $\mu_2(X_k)>0$ for at least one k $(k=1, 2, \cdots, n)$. The non-negative square root of $\mu_2(X_k)$ is the standard deviation of X_k and will be denoted by σ_k. We call

$$(1)\qquad \lambda(X_k) = \begin{cases} \mu_3(X_k)/\mu_2(X_k), & \text{if } \mu_2(X_k) \neq 0, \\ 0, & \text{if } \mu_2(X_k) = 0, \end{cases}$$

the *moment-ratio* of X_k. If X_k is essentially bounded(5) and its bound, meas-

Presented to the Society, February 24, 1940; received by the editors May 8, 1940.

(1) H. Cramér, *Random Variables and Probability Distributions* (Cambridge Tracts in Mathematics, no. 36), Cambridge University Press, 1937, pp. 56–64.

(2) A. Liapounoff, *Nouvelle forme du théorème sur la limite de probabilité*, Mémoires de l'Académie des Sciences de St-Petersbourg, (8), vol. 12 (1901). (Cramér, loc. cit., p. 60.)

(3) J. W. Lindeberg, *Eine neue Herleitung des Exponentialgesetzes in der Wahrscheinlichkeitsrechnung*, Mathematische Zeitschrift, vol. 15 (1922), pp. 211–225. (Cramér, loc. cit., Theorem 21.)

(4) W. Feller, *Über den zentralen Grenzwertsatz der Wahrscheinlichkeitsrechnung*, Mathematische Zeitschrift, vol. 40 (1935), pp. 521–559. (Cramér, loc. cit., Theorem 22.)

(5) The author originally developed the early sections of this paper for the case of bounded variates, and is indebted to W. Feller who urged the study, in these sections, of the case of finite third order absolute moments.

ured from its mean value, is denoted by $l(X_k)$, then, as can be verified easily,

$$\lambda(X_k) \leqq l(X_k). \tag{2}$$

We set

$$\Lambda = \max \{\lambda(X_1), \lambda(X_2), \cdots, \lambda(X_n)\}. \tag{3}$$

The familiar inequality(6) that the square root of the second order absolute moment cannot exceed the cube root of the third order absolute moment readily yields the set of estimates:

$$\sigma_k \leqq \Lambda, \qquad (k = 1, 2, \cdots, n). \tag{4}$$

Consider, now, the variate sum

$$X = X_1 + X_2 + \cdots + X_n. \tag{5}$$

It has the mean value

$$\alpha = \alpha_1 + \alpha_2 + \cdots + \alpha_n, \tag{6}$$

and, in virtue of the independence of the given variates, the standard deviation

$$\sigma = (\sigma_1^2 + \sigma_2^2 + \cdots + \sigma_n^2)^{1/2}. \tag{7}$$

The formula which we are going to derive involves the two numbers Λ and σ, respectively the maximum moment-ratio of the *individual* variates and the standard deviation of their *sum*. Our assumption that the variates are not all constant implies that these numbers are both positive. We introduce their ratio

$$\epsilon = \Lambda/\sigma. \tag{8}$$

It is this ratio which serves as a convenient measure of the extent to which the sum X fails to be Gaussian.

The Gaussian (or Laplacean) distribution function

$$G(x) = \frac{1}{(2\pi)^{1/2}} \int_{-\infty}^{x} e^{-t^2/2} dt, \qquad (-\infty < x < \infty), \tag{9}$$

characterizes the law of probability obeyed by a normal variate of mean value 0 and standard deviation 1. For every x, $G(x)$ is the probability that such a variate assume a value less than x. A normal variate of mean value α and standard deviation σ has the distribution function $G((x-\alpha)/\sigma)$. Let the variate sum X have the distribution function $F(x)$. The least upper bound,

$$M = \sup_{-\infty < x < \infty} \left| F(x) - G((x - \alpha)/\sigma) \right|, \tag{10}$$

(6) Hardy-Littlewood-Pólya, *Inequalities*, Cambridge, 1934, p. 157.

of the modulus of the difference between $F(x)$ and the associated normal distribution function constitutes a precise measure of the "abnormality" of X. A theorem of Liapounoff, about which we shall have more to say in §4, implies

$$\lim_{\epsilon \to 0} M = 0. \tag{11}$$

Our first goal, an arithmetical refinement of (11), reads:

THEOREM 1. $M \leqq (1.88)\epsilon$.

We postpone the proof. The theorem shows that the ratios M/ϵ arising from admissible sets of variates constitute a bounded aggregate of real numbers. Let

$$C = \sup M/\epsilon. \tag{12}$$

Then, for each admissible set of variates, it will be true that $M \leqq C\epsilon$. In addition, corresponding to each positive $C' < C$ there will exist an admissible set of variates having $M > C'\epsilon$. By Theorem 1, $C \leqq 1.88$. We shall now prove the following

THEOREM 2. $C \geqq 1/(2\pi)^{1/2}$.

Proof. Let n be an odd positive integer. Let $X_1, X_2, \cdots, X_n$ be totally independent but similar. Indeed, let each X_k have only the two possible values $+1$ and -1, each with the associated probability 1/2. We see that $\epsilon = 1/n^{1/2}$. The variate sum X has the $(n+1)$ possible values $\pm 1, \pm 3, \cdots, \pm n$. Its distribution function $F(x)$ is a step-function which is constant in each interval free of these possible values. By symmetry, $F(x) = 1/2$ throughout $-1 < x < 1$. Clearly,

$$M \geqq \lim_{x \to 1\ (x<1)} |F(x) - G(\epsilon x)| = G(\epsilon) - G(0).$$

If we assume $C < 1/(2\pi)^{1/2}$ we can determine n correspondingly large so as to force

$$\frac{M}{\epsilon} \geqq \frac{G(\epsilon) - G(0)}{\epsilon} > C.$$

This, because

$$\lim_{\epsilon \to 0} \frac{G(\epsilon) - G(0)}{\epsilon} = G'(0) = \frac{1}{(2\pi)^{1/2}} \cdot$$

But the variates under consideration form an admissible set and so have $M \leqq C\epsilon$. The contradiction establishes the theorem.

Remark. If, for a given $\eta > 0$, we denote by C_η the least upper bound of those ratios M/ϵ which arise from admissible sets of variates having $\epsilon < \eta$, it is clear that C_η is a monotone increasing function of η. In particular, $C_\eta \leqq C$

for all η. Since, in the proof of Theorem 2, we may replace a satisfactory n by any larger odd integer, we infer that $C_\eta \geqq 1/(2\pi)^{1/2}$ for all η.

Reduction to the case (R). We now take advantage of the fact that moments have been measured about mean values, the fact that Λ and σ enter Theorem 1 only in their ratio ϵ, and our claim that $C \leqq 1.88$. We shall say that we have the case (R) if the given variates meet the additional requirements:

$$\alpha_k = 0, \qquad (k = 1, 2, \cdots, n), \tag{13}$$

$$\sigma = 1, \tag{14}$$

$$\epsilon < \frac{1}{1.88}. \tag{15}$$

The corresponding case of Theorem 1 we shall call Theorem 1(R). We are going to show that the general theorem is a corollary of its own special case. Noting that distribution functions assume only values between 0 and 1, hence that the inequality $M \leqq 1$ is always valid, we see that Theorem 1 is trivial when (15) is false. We may confine our attention, therefore, to the case (15). If, however, the original variates of Theorem 1 do not also satisfy (13) and (14), we introduce the associated variates

$$X_k' = \frac{1}{\sigma}(X_k - \alpha_k), \qquad (k = 1, 2, \cdots, n).$$

These have $M' = M$, $\epsilon' = \epsilon$, and satisfy in detail the hypotheses of Theorem 1(R), all of which can be demonstrated without difficulty. Since the inequalities $M' \leqq (1.88)\epsilon'$ and $M \leqq (1.88)\epsilon$ are equivalent, Theorem 1 is a consequence of Theorem 1(R).

Elementary properties of $F(x) - G(x)$. In the present case (R), the sum X is a reduced variate: its mean value is 0 and its standard deviation is 1. The Bienaymé-Tchebycheff(7) inequality for a reduced variate reads:

$$\Pr\{|X| \geqq x\} \leqq 1/x^2, \qquad (x > 0).$$

Interpreting this in terms of $F(x)$, and equally well for $G(x)$, we infer

$$|F(x) - G(x)| \leqq 1/x^2, \qquad (-\infty < x < \infty). \tag{16}$$

It follows that a sequence of points x on which the modulus $|F(x) - G(x)|$ tends to its least upper bound M forms a bounded set. An easily constructed argument establishes the existence of a finite point $x = b$ for which either

$$F(b+) - G(b) = M, \tag{17}$$

or

$$F(b-) - G(b) = -M. \tag{17'}$$

(7) Cramér, loc. cit., p. 21.

LEMMA 1. *There exists a number a such that one of the inequalities*

$$F(x+a)-G(x+a) \geqq \frac{\delta - x}{(2\pi)^{1/2}}, \tag{18}$$

$$F(x+a)-G(x+a) \leqq \frac{-\delta - x}{(2\pi)^{1/2}} \tag{18'}$$

holds throughout the interval $-\delta<x<\delta$ *where*

$$\delta = M\cdot\left(\frac{\pi}{2}\right)^{1/2}. \tag{19}$$

Proof. In case (17) write $a=b+\delta$. Then, for all $x>b$,

$$F(x) \geqq F(b+), \qquad G(x) \leqq G(b)+\frac{x-b}{(2\pi)^{1/2}},$$

the first because $F(x)$ is non-decreasing, the second since $G(x)$ is differentiable and $G'(x) \leqq 1/(2\pi)^{1/2}$. Whence,

$$F(x)-G(x) \geqq M - \frac{x-b}{(2\pi)^{1/2}}$$

for all $x>b$. This implies (18) for all $x>-\delta$, *a fortiori* for $-\delta<x<\delta$. In case (17′) a symmetric argument employing $a=b-\delta$ establishes (18′) in $-\delta<x<\delta$.

2. The proof of Theorem 1(R)

Our proof, which rests on the calculation to be presented in §3, utilizes the characteristic functions

$$\phi(t) = \int_{-\infty}^{\infty} e^{ixt}dF(x), \tag{20}$$

$$(-\infty < t < \infty),$$

$$\psi(t) = \int_{-\infty}^{\infty} e^{ixt}dG(x) = e^{-t^2/2}, \tag{21}$$

of the distributions under discussion. In (21) the indicated evaluation is familiar. Since $G(x)$ is a differentiable function, the Stieltjes integral can be written as an ordinary integral:

$$\psi(t) = \int_{-\infty}^{\infty} e^{ixt}G'(x)dx = \frac{1}{(2\pi)^{1/2}}\int_{-\infty}^{\infty} e^{-x^2/2}e^{ixt}dx.$$

Thus, $\psi(t)$ is the Fourier transform of $e^{-x^2/2}$. That this is $e^{-t^2/2}$ can be verified readily.

We are interested primarily in the difference

$$\phi(t) - \psi(t) = \int_{-\infty}^{\infty} e^{ixt} d\{F(x) - G(x)\}. \tag{22}$$

From (16) we infer that $F(x)-G(x)$ is absolutely integrable over $-\infty<x<\infty$ and tends to 0 when $x\to\pm\infty$. These facts justify the integration by parts which yields

$$\frac{\phi(t) - \psi(t)}{-it} = \int_{-\infty}^{\infty} \{F(x) - G(x)\} e^{ixt} dx. \tag{23}$$

It is known([8]) that if the difference (22) is uniformly small in some finite interval about $t=0$, then M is small. This fact suggests the following procedure.

In (23) we replace x by $x+a$ (the a of Lemma 1) and obtain

$$\frac{\phi(t) - \psi(t)}{-it} \cdot e^{-iat} = \int_{-\infty}^{\infty} \{F(x+a) - G(x+a)\} e^{ixt} dx. \tag{24}$$

We confine t to the finite interval

$$-T \leqq t \leqq T, \tag{25}$$

by employing a weighting factor $w(t)$ which vanishes outside this interval. We choose

$$w(t) = T - |t|, \qquad \text{when } -T \leqq t \leqq T. \tag{26}$$

The Fourier transform of $w(t)$ is, except for a missing constant factor,

$$W(x) = \int_{-T}^{T} w(t) e^{ixt} dt = \frac{2(1-\cos Tx)}{x^2}. \tag{27}$$

The Parseval theorem in the theory of Fourier transforms assures us of the validity of the equality

$$\int_{-\infty}^{\infty} W(x)\{F(x+a) - G(x+a)\} dx = \int_{-T}^{T} w(t) \frac{\phi(t) - \psi(t)}{-it} e^{-iat} dt. \tag{28}$$

This may be derived directly by multiplying (24) throughout by $w(t)$, integrating over (25), and inverting the order of integration in the resulting iterated integral. The last step is justified by the absolute integrability of the product $w(t)\{F(x+a)-G(x+a)\}$ over the strip $-T\leqq t\leqq T$, $-\infty<x<\infty$.

Since, as a brief inspection of (22) shows, the modulus $|\phi(t)-\psi(t)|$ is an even function of t, we can derive from the Parseval equality (28) the inequality

---

([8]) P. Lévy, *Théorie de l'Addition des Variables Aléatoires* (Monographies des Probabilités), Paris, 1937, p. 49. Cramér, loc. cit., p. 29, Theorem 11.

$$\left|\int_{-\infty}^{\infty} \frac{1-\cos Tx}{x^2}\{F(x+a)-G(x+a)\}dx\right| \tag{29}$$
$$\leqq \int_0^T (T-t)\frac{|\phi(t)-\psi(t)|}{t}dt.$$

Noting that $|F(x)-G(x)| \leqq M = \delta(2/\pi)^{1/2}$ for all x, we find

$$\left|\left(\int_{-\infty}^{-\delta}+\int_{\delta}^{\infty}\right)\frac{1-\cos Tx}{x^2}\{F(x+a)-G(x+a)\}dx\right|$$
$$\leqq 2(2/\pi)^{1/2}\delta\int_{\delta}^{\infty}\frac{1-\cos Tx}{x^2}dx.$$

On the other hand, Lemma 1 yields, equally in its two cases,

$$\left|\int_{-\delta}^{\delta}\frac{1-\cos Tx}{x^2}\{F(x+a)-G(x+a)\}dx\right| \geqq (2/\pi)^{1/2}\delta\int_0^{\delta}\frac{1-\cos Tx}{x^2}dx.$$

These two estimates, the standard evaluation

$$\int_0^{\infty}\frac{1-\cos x}{x^2}dx = \frac{\pi}{2},$$

the triangle inequality

$$\left|\int_{-\infty}^{\infty}\right| \geqq \left|\int_{-\delta}^{\delta}\right| - \left|\int_{-\infty}^{-\delta}+\int_{\delta}^{\infty}\right|,$$

and certain obvious reductions enable us to deduce from (29) the following:

$$A(T\delta) \leqq \int_0^T (T-t)\frac{|\phi(t)-\psi(t)|}{t}dt, \tag{30}$$

where

$$A(u) = (2/\pi)^{1/2}\cdot u\cdot\left\{3\int_0^u \frac{1-\cos x}{x^2}dx - \pi\right\}. \tag{31}$$

In §3, for the particular choice $T=1.1/\epsilon$, we shall prove that the right member of (30) does not exceed

$$\frac{1.1}{6}\cdot\left(\frac{\pi}{2}\right)^{1/2}.$$

Thus,

$$A\left(\frac{1.1}{\epsilon}\delta\right) < 0.2298. \tag{32}$$

Granting this for the moment, we employ the Taylor inequality

$$\cos x \leqq 1 - \frac{x^2}{2!} + \frac{x^4}{4!} - \frac{x^6}{6!} + \frac{x^8}{8!},$$

and make the calculation:

$$A(2.59) > 0.2298. \tag{33}$$

Now, at least when it is positive, $A(u)$ is an increasing function. It follows that

$$M = (2/\pi)^{1/2}\cdot\delta \leqq (2/\pi)^{1/2}\cdot\frac{(2.59)\epsilon}{1.1} < (1.88)\epsilon. \tag{34}$$

We have completed the proof of the fact that Theorem 1(R) rests on the calculation to be presented in the next section.

3. The calculation

The individual distribution functions. We must return to the individual variates $X_1, X_2, \cdots, X_n$ of Theorem 1(R). The corresponding distribution functions $F_1(x), F_2(x), \cdots, F_n(x)$ all increase monotonically and have the common limits 0, 1 when x extends respectively to $-\infty, +\infty$. The hypotheses of Theorem 1(R) require, for each k $(k=1, 2, \cdots, n)$,

$$\int_{-\infty}^{\infty} x dF_k(x) = 0, \qquad \int_{-\infty}^{\infty} x^2 dF_k(x) = \sigma_k^2,$$

$$\int_{-\infty}^{\infty} |x|^3 dF_k(x) = \mu_3(X_k) \leqq \epsilon\cdot\sigma_k^2.$$

Inequalities (4) here read: $\sigma_k \leqq \epsilon$. We shall feel free to use the foregoing facts without further explicit reference.

The individual characteristic functions. The variate X_k has the characteristic function

$$\phi_k(t) = \int_{-\infty}^{\infty} e^{ixt} dF_k(x) = 1 - \tfrac{1}{2}\sigma_k^2 t^2 + \cdots.$$

The constant term and, better, the two indicated terms of the formal Taylor expansion will be employed as approximations. We shall need estimates for the errors

$$u_k = \phi_k(t) - 1 = \int_{-\infty}^{\infty} (e^{ixt} - 1 - ixt) dF_k(x), \tag{35}$$

$$v_k = \phi_k(t) - 1 + \tfrac{1}{2}\sigma_k^2 t^2 = \int_{-\infty}^{\infty} (e^{ixt} - 1 - ixt + \tfrac{1}{2}x^2t^2) dF_k(x). \tag{36}$$

The elementary inequalities

$$|e^{ixt}-1-ixt| \leqq \tfrac{1}{2}x^2t^2, \quad |e^{ixt}-1-ixt+\tfrac{1}{2}x^2t^2| \leqq \tfrac{1}{6}|x|^3|t|^3,$$

which are valid for all real x and t, at once yield

$$|u_k| \leqq \tfrac{1}{2}\sigma_k^2 t^2, \tag{37}$$

$$|v_k| \leqq \tfrac{1}{6}\epsilon\cdot\sigma_k^2|t|^3. \tag{38}$$

We now confine our attention to the interval

$$0 \leqq t < \frac{2^{1/2}}{\epsilon}, \tag{39}$$

in which, by virtue of (37), we may be certain that $|u_k|<1$. In this interval, then, $\phi_k(t)$ is never 0 and we may define

$$\log \phi_k(t) = \log(1+u_k) = \int_0^{u_k}\frac{dz}{1+z}, \tag{40}$$

for example by employing as the path of integration in the complex plane the straight line segment joining $z=0$ to $z=u_k$. For the difference

$$r_k = \log(1+u_k) - u_k = -\int_0^{u_k}\frac{zdz}{1+z},$$

we obtain the estimate

$$|r_k| \leqq \int_0^{|u_k|}\frac{xdx}{1-x} \leqq \int_0^{\sigma_k^2t^2/2}\frac{xdx}{1-x} = \sigma_k^4\int_0^{t^2/2}\frac{ydy}{1-\sigma_k^2 y}$$

$$\leqq \epsilon^2\sigma_k^2\int_0^{t^2/2}\frac{ydy}{1-\epsilon^2 y} = -\sigma_k^2t^2\left\{\frac{\log(1-\frac{1}{2}\epsilon^2t^2)+\frac{1}{2}\epsilon^2t^2}{\epsilon^2t^2}\right\}.$$

If we write

$$h(x) = \frac{x}{6} - \frac{1}{x^2}\left\{\log\left(1-\frac{x^2}{2}\right)+\frac{x^2}{2}\right\}, \qquad (0 \leqq x < 2^{1/2}), \tag{41}$$

and observe that

$$\log \phi_k(t) = -\tfrac{1}{2}\sigma_k^2t^2 + v_k + r_k,$$

we find that we have proved

LEMMA 2. *Throughout the interval* (39),

$$\log \phi_k(t) = -\tfrac{1}{2}\sigma_k^2t^2 + \sigma_k^2\Delta_k,$$

where, for each k $(k=1, 2, \cdots, n)$, $|\Delta_k| \leqq t^2\cdot h(\epsilon t)$.

The characteristic function of the variate sum. We now encounter the essential reason for the hypothesis that the given variates be totally independent. This hypothesis guarantees the relation

$$\phi(t) = \phi_1(t)\phi_2(t) \cdots \phi_n(t). \tag{42}$$

This implication is well known[9]. With a suitable determination of the logarithm of the product, we can now derive from Lemma 2

LEMMA 3. *Throughout the interval* (39),

$$\log \phi(t) = -\tfrac{1}{2}t^2 + \Delta,$$

where

$$|\Delta| \leqq t^2 \cdot h(\epsilon t).$$

Recalling that $\psi(t) = e^{-t^2/2}$, and observing that

$$\phi(t) - \psi(t) = (e^{\Delta} - 1)\psi(t),$$
$$|e^{\Delta} - 1| \leqq e^{|\Delta|} - 1,$$

we obtain from Lemma 3 the final

LEMMA 4. *Throughout the interval* (39),

$$|\phi(t) - \psi(t)| \leqq \{e^{t^2 \cdot h(\epsilon t)} - 1\}e^{-t^2/2}.$$

The integral B. We are now in position to calculate an upper bound for the integral which constitutes the right member of the inequality (30). For the choice $T = 1.1/\epsilon$ it is clear from Lemma 4 that this integral is dominated in magnitude by the integral

$$B = \int_0^{1.1/\epsilon} \left(\frac{1.1}{\epsilon} - t\right) \frac{e^{t^2 \cdot h(\epsilon t)} - 1}{t} e^{-t^2/2} dt. \tag{43}$$

We shall divide the calculation into three major parts by writing

$$B = B_1 + B_2 + B_3, \tag{44}$$

where, in terms of the abbreviation $c = 0.75$,

$$B_1 = \frac{\epsilon}{5} \int_0^{c/\epsilon} \left(\frac{1.1}{\epsilon} - t\right) t^2 e^{-t^2/2} dt,$$

$$B_2 = \int_0^{c/\epsilon} \left(\frac{1.1}{\epsilon} - t\right) \frac{e^{t^2 \cdot h(\epsilon t)} - 1 - \frac{1}{5}\epsilon t^3}{t} e^{-t^2/2} dt, \tag{45}$$

$$B_3 = \int_{c/\epsilon}^{1.1/\epsilon} \left(\frac{1.1}{\epsilon} - t\right) \frac{e^{t^2 \cdot h(\epsilon t)} - 1}{t} e^{-t^2/2} dt.$$

---

(9) Cramér, loc. cit., p. 36.

The term B_1. With the aid of a few integrations by parts, and standard evaluations, we find we can write B_1 in the form

$$B_1 = \frac{1.1}{6}\left(\frac{\pi}{2}\right)^{1/2} - \frac{\epsilon}{3} - \frac{1}{6}\int_{c/\epsilon}^{\infty}\left\{(1.1 - c)t^2 + c - \frac{2c}{t^2}\right\}e^{-t^2/2}dt.$$

The factor in braces in the integrand is an increasing function of t. At precisely this point we use the hypothesis (15). This implies that $t > (1.88)c$ in the mentioned factor. We infer that this factor is positive. Hence,

$$B_1 \leqq \frac{1.1}{6}\left(\frac{\pi}{2}\right)^{1/2} - \frac{\epsilon}{3}. \tag{46}$$

The term B_2. We begin with the observation that, for each fixed positive $t \leqq c/\epsilon$,

$$\frac{1}{\epsilon^2}\left(e^{t^2\cdot h(\epsilon t)} - 1 - \frac{\epsilon t^3}{6}\right)$$

is an increasing function of ϵ throughout $0 < \epsilon \leqq c/t$. This can be proved, for example, by differentiating with respect to ϵ and employing $1 \leqq e^{t^2\cdot h(\epsilon t)}$ and $\frac{1}{6} = h'(0) \leqq h'(\epsilon t)$, to show that this derivative is positive. It follows that

$$B_2 \leqq B_2' = \frac{\epsilon^2}{c^2}\int_0^{c/\epsilon}\left(\frac{1.1}{\epsilon} - t\right)t\left\{e^{t^2\cdot h(c)} - 1 - \frac{ct^2}{6}\right\}e^{-t^2/2}dt.$$

Let us increase the upper limit of integration from c/ϵ to $1.1/\epsilon$ thus replacing B_2' by a still larger number, say B_2''. An integration by parts then shows that

$$B_2'' = \frac{(1.1)\epsilon}{c^2}\left\{\frac{1}{[1 - 2h(c)]^{1/2}} - 1 - \frac{c}{3}\right\}$$
$$- \frac{\epsilon}{c}\int_0^{1.1/\epsilon}\left\{\frac{e^{t^2\cdot h(c)}}{[1 - 2h(c)]^{1/2}} - 1 - \frac{c}{3} - \frac{ct^2}{6}\right\}e^{-t^2/2}dt.$$

The computation $0.2121 < h(c) < 0.2122$ proves $c[1 - 2h(c)]^{1/2} < 6h(c)$ and so demonstrates that the factor in braces in the last integrand is an increasing function of t^2. Since this factor is positive at $t = 0$, it follows that we may neglect the integral. Thus,

$$B_2 \leqq (0.134)\epsilon. \tag{47}$$

The term B_3. From the numerator in the middle factor of the integrand in B_3 we reject the term -1. The resulting larger integral we call B_3'. This can be written

$$B_3' = \frac{1}{\epsilon}\int_{\epsilon}^{1.1}\left(\frac{1.1}{u} - 1\right)e^{-H(u)/\epsilon^2}du,$$

where

$$H(u) = u^2\{\tfrac{1}{2} - h(u)\}.$$

It can be seen that $H(u) > 0$ in the interval of integration. This permits us to employ the elementary inequality

$$x \cdot e^{-x} \leqq 1/e, \qquad (0 < x),$$

and find

$$B_3' \leqq \frac{\epsilon}{e}\int_c^{1.1}\left(\frac{1.1}{u} - 1\right)\frac{du}{H(u)}\cdot$$

An examination of the derivative

$$H'(u) = u\cdot\left(1 - \frac{u}{2} - \frac{u^2}{2-u^2}\right)$$

shows that $H(u)$ increases in $c \leqq u \leqq u'$ and decreases in $u' \leqq u \leqq 1.1$ where u' is the unique root of $H'(u)$ in the interval of integration. We find $0.85 < u' < 0.851$. Since $H(1) < H(0.75)$ we infer that

$$H(u) \geqq \begin{cases} H(1) & > 0.14010 \text{ when } 0.75 \leqq u \leqq 1, \\ H(1.05) & > 0.10814 \text{ when } 1 \leqq u \leqq 1.05, \\ H(1.1) & > 0.05674 \text{ when } 1.05 \leqq u \leqq 1.1. \end{cases}$$

By dividing the interval of integration into the three indicated parts and by replacing the reciprocal of $H(u)$ in each part by its upper bound in that part, we obtain the estimate

$$B_3 \leqq (0.195)\epsilon. \tag{48}$$

Summary of calculation. We have proved

$$B \leqq \frac{1.1}{6}\left(\frac{\pi}{2}\right)^{1/2} + \epsilon(-\tfrac{1}{3} + 0.134 + 0.195) < \frac{1.1}{6}\left(\frac{\pi}{2}\right)^{1/2}. \tag{49}$$

This shows that (32) is a valid consequence of (30) and so completes the argument establishing Theorem 1(R) and its dependent extension Theorem 1.

4. Generalizations

Let $X_1, X_2, \cdots, X_n$ be totally independent variates with the respective distribution functions $F_1(x), F_2(x), \cdots, F_n(x)$. Let $s > 0$ and $a_1, a_2, \cdots, a_n$ be any real numbers. We wish to compare the distribution function $F(x)$ of the variate sum $X = X_1 + X_2 + \cdots + X_n$ with the distribution function $G((x-a)/s)$ of a normal variate of mean value $a = a_1 + a_2 + \cdots + a_n$ and standard deviation s. We put

$$M = \sup_{-\infty < x < \infty} \left| F(x) - G((x-a)/s) \right|.$$

For a given $\epsilon > 0$ we introduce the three quantities

$$\epsilon_0 = \sum_{k=1}^{n} \Pr\left\{ \left| X_k - a_k \right| > \epsilon s \right\},$$

$$\epsilon_1 = \frac{1}{s} \cdot \sum_{k=1}^{n} \left| \int_{a_k - \epsilon s}^{a_k + \epsilon s} (x_k - a_k) dF_k(x) \right|,$$

$$\epsilon_2 = \left| 1 - \frac{1}{s^2} \cdot \sum_{k=1}^{n} \int_{a_k - \epsilon s}^{a_k + \epsilon s} (x_k - a_k)^2 dF_k(x) \right|.$$

Feller (loc. cit.) has proved that if these three numbers are small then M is small. We shall now establish

THEOREM 3. *If $\epsilon_0 \leqq \epsilon$, $\epsilon_1 \leqq \epsilon$, $\epsilon_2 \leqq \epsilon$, then*

$$M \leqq (5.8)\epsilon.$$

More generally, whenever $\epsilon_1^2 + \epsilon_2 < 1$, we have

$$M \leqq \frac{C(\epsilon + \epsilon_1)}{(1 - \epsilon_1^2 - \epsilon_2)^{1/2}} + \epsilon_0 + \frac{\epsilon_1}{(2\pi)^{1/2}} + \frac{1}{(2\pi e)^{1/2}} \log \frac{1}{(1 - \epsilon_1^2 - \epsilon_2)^{1/2}},$$

where C ($\leqq 1.88$) is the constant furnished by Theorem 1.

Proof. Without loss of generality we may confine our attention to the special case

$$\text{(R)} \qquad s = 1, \; a_1 = a_2 = \cdots = a_n = 0.$$

(We have but to introduce $X_k' = (1/s)(X_k - a_k)$ and to set $s' = 1$, $a_1' = a_2' = \cdots = a_n' = 0$, $\epsilon' = \epsilon$ in order to discover, first, that $M' = M$, $\epsilon_0' = \epsilon_0$, $\epsilon_1' = \epsilon_1$, $\epsilon_2' = \epsilon_2$ and, therefore, that the theorem is equivalent to its special case.)

We use a familiar device[10]. We approximate the given variates by the associated bounded variates

$$\overline{X}_k = \begin{cases} X_k, & \text{if } \left| X_k \right| \leqq \epsilon, \\ 0, & \text{if } \left| X_k \right| > \epsilon, \end{cases} \qquad (k = 1, 2, \cdots, n).$$

To these totally independent variates we are going to apply Theorem 1. Now, $\overline{X}_k$ has the mean value

$$\bar{a}_k = \int_{-\epsilon}^{\epsilon} x dF_k(x), \qquad (k = 1, 2, \cdots, n).$$

This is easily verified. Next, since

(10) Lévy, loc. cit., pp. 104–110.

$$\sum_{k=1}^{n} |\bar{\alpha}_k| = \epsilon_1, \tag{50}$$

we see that the bound of $\bar{X}_k$, measured from the mean value, does not exceed $\epsilon+\epsilon_1$. Recalling (2) we find

$$\bar{\Lambda} \leqq \epsilon + \epsilon_1, \tag{51}$$

where $\bar{\Lambda}$ is the maximum moment-ratio of the variates $\bar{X}_k$. The sum, $\bar{X}$, of these variates has the standard deviation $\bar{\sigma}$ given by

$$\bar{\sigma}^2 = \sum_{k=1}^{n} \int_{-\epsilon}^{\epsilon} x^2 dF_k(x) - \sum_{k=1}^{n} \bar{\alpha}_k^2.$$

The second term of this difference is non-negative and, by (50), not greater than ϵ_1^2. Thus,

$$(1 - \epsilon_1^2 - \epsilon_2)^{1/2} \leqq \bar{\sigma} \leqq (1 + \epsilon_2)^{1/2}. \tag{52}$$

By Theorem 1,

$$\left| \bar{F}(x) - G\left(\frac{x - \bar{\alpha}}{\bar{\sigma}}\right) \right| \leqq \frac{C(\epsilon + \epsilon_1)}{(1 - \epsilon_1^2 - \epsilon_2)^{1/2}}, \qquad (-\infty < x < \infty), \tag{53}$$

where $\bar{F}(x)$ denotes the distribution function of $\bar{X}$, and $\bar{\alpha}$ its mean value.

Next, we observe that $\bar{X}$ differs from X at most in those cases in which for at least one k, $|X_k| > \epsilon$. These cases have a total probability of occurrence not greater than ϵ_0. Hence,

$$|F(x) - \bar{F}(x)| \leqq \epsilon_0, \qquad (-\infty < x < \infty). \tag{54}$$

If we combine (53) and (54) we obtain an inequality sharper, in some respects, than that announced in the theorem. But we must pass from $G((x-\bar{\alpha})/\bar{\sigma})$ to the reduced normal distribution function $G(x)$ desired in the present case (R). Since, by (50), $|\bar{\alpha}| \leqq \epsilon_1$, and since $e^{-t^2/2} \leqq 1$, we have

$$|G(x) - G(x - \bar{\alpha})| \leqq \frac{\epsilon_1}{(2\pi)^{1/2}}, \qquad (-\infty < x < \infty). \tag{55}$$

And the elementary inequality $|te^{-t^2/2}| \leqq 1/e^{1/2}$ yields, for all x,

$$\left| G(x - \bar{\alpha}) - G\left(\frac{x - \bar{\alpha}}{\bar{\sigma}}\right) \right| \leqq \frac{|\log \bar{\sigma}|}{(2\pi e)^{1/2}} \leqq \frac{1}{(2\pi e)^{1/2}} \log \frac{1}{(1 - \epsilon_1^2 - \epsilon_2)^{1/2}}. \tag{56}$$

The general inequality of the theorem is an immediate consequence of the inequalities (53)–(56).

Finally, we consider the special case $\epsilon_0 \leqq \epsilon$, $\epsilon_1 \leqq \epsilon$, $\epsilon_2 \leqq \epsilon$. Since it is always true that $M \leqq 1$ we may assume that $\epsilon < 1/5.8$, the desired inequality $M \leqq (5.8)\epsilon$ being trivial in the contrary case. Thus, the condition $\epsilon_1^2 + \epsilon_2 < 1$

is met generously and we may apply the already established general inequality. This yields

$$\frac{M}{\epsilon} \leqq \frac{3.76}{(1-\epsilon^2-\epsilon)^{1/2}} + 1 + \frac{1}{(2\pi)^{1/2}} + \frac{1}{\epsilon(2\pi e)^{1/2}} \log \frac{1}{(1-\epsilon^2-\epsilon)^{1/2}}.$$

It is easily seen that the right member is an increasing function of ϵ in $0<\epsilon<1/5.8$ and is less than 5.8 when $\epsilon=1/5.8$. This proves the theorem.

The case (L). Let the standard deviation of each X_k be finite. Let $s=\sigma$, $a_1=\alpha_1$, $a_2=\alpha_2, \cdots, a_n=\alpha_n$, where the α's are the mean values of the individual variates and σ is the standard deviation of their sum. If these conditions are met we shall say we have the case (L). In this case we can express ϵ_1 and ϵ_2 in the convenient forms:

$$\epsilon_1 = \frac{1}{\sigma}\cdot \sum_{k=1}^{n} \left| \left(\int_{-\infty}^{\alpha_k-\epsilon\sigma} + \int_{\alpha_k+\epsilon\sigma}^{\infty} \right)(x-\alpha_k)dF_k(x) \right|,$$

$$\epsilon_2 = \frac{1}{\sigma^2}\cdot \sum_{k=1}^{n} \left(\int_{-\infty}^{\alpha_k-\epsilon\sigma} + \int_{\alpha_k+\epsilon\sigma}^{\infty} \right)(x-\alpha_k)^2dF_k(x).$$

These imply

$$\epsilon_2 \geqq \epsilon^2\epsilon_0, \qquad \epsilon_2 \geqq \epsilon\cdot\epsilon_1. \tag{57}$$

Lindeberg (loc. cit.) has proved that if ϵ_2 is small then M is small. We shall now establish

THEOREM 4. *In case* (L) *if* $\epsilon_2 \leqq \epsilon^3$, *then* $M \leqq (3.6)\epsilon$.

Proof. We shall assume $\epsilon<1/3.6$ since the theorem is trivial in the contrary case. By (57), we have $\epsilon_0 \leqq \epsilon$, $\epsilon_1 \leqq \epsilon^2$. The general inequality of Theorem 3 yields

$$\frac{M}{\epsilon} \leqq \frac{(1.88)(1+\epsilon)}{(1-\epsilon^4-\epsilon^3)^{1/2}} + 1 + \frac{1}{(2\pi)^{1/2}} + \frac{1}{\epsilon(2\pi e)^{1/2}} \log \frac{1}{(1-\epsilon^4-\epsilon^3)^{1/2}}.$$

The right member is an increasing function of ϵ in $0<\epsilon<1/3.6$ and has a value <3.6 when $\epsilon=1/3.6$. This proves the theorem.

Finally, we consider the subcase in which for some (not necessarily integral) $m>2$, each X_k has a finite absolute moment (about its mean value) of order m: $\mu_m(X_k)$. We put $\eta=(1/\sigma^m)\sum_{k=1}^{n}\mu_m(X_k)$. Liapounoff (loc. cit.) proved M small when η is small.

THEOREM 5. $M \leqq (3.6)\eta^{1/(m+1)}$.

Proof. We define $\epsilon=\eta^{1/(m+1)}$. For this ϵ it is easy to prove $\epsilon_2 \leqq \epsilon^3$. The present theorem thus follows from Theorem 4.

COLUMBIA UNIVERSITY,
NEW YORK, N. Y.

MINIMAL POSITIVE HARMONIC FUNCTIONS

BY

ROBERT S. MARTIN

Introduction. One may ask how great generality in a domain is to be permitted if we are to have for this domain a formula possessing the more significant features of the Poisson-Stieltjes integral formula for the circle or the sphere(1). Even if there is agreement as to what the more important consequences of the formula are, there are two approaches, differing not so much in content as in emphasis, along which partial answers to the question lie. The first consists in determining hypotheses, as weak as possible, upon a domain under which all, or substantially all, of the important features of the formula admit of extension. The second consists in attempting to determine for each of the important consequences of the formula the class of domains for which it holds. While this sounds very much like the distinction between obtaining sufficient and obtaining necessary conditions for an extension of the formula together with all of its important features, actually it goes a little deeper, since the second viewpoint involves implicitly the notion that what a significant extension of the formula is may depend upon what it is going to be used for. It is a particular consideration from the viewpoint of the second approach which leads to the concept of a minimal positive harmonic function with which we are concerned in the present article.

A function positive and harmonic in a given domain we shall call *minimal*(2)—for this domain—if it dominates there no positive harmonic function except for its own constant submultiples. An important instance of this kind of function occurs in connection with the principle of Picard(3), whose relation

---

Presented to the Society, October 26, 1940; received by the editors May 10, 1940.

(1) For the two-dimensional case the possibilities have been rather fully discussed by G. C. Evans, *The Logarithmic Potential*, American Mathematical Society Colloquium Publications, vol. 6, New York, 1927, esp. chaps. 5 and 6. For the three-dimensional case under hypotheses related to bounded curvature of the boundaries, see C. de la Vallée Poussin, *Propriétés des fonctions harmoniques dans un domaine ouvert limité par des surfaces à courbure bornée*, Annali della R. Scuola Normale Superiore di Pisa, (2), vol. 2 (1933), pp. 167–192; George A. Garrett, *Necessary and sufficient conditions for potentials of single and double layers*, American Journal of Mathematics, vol. 58 (1936), pp. 95–129. An approach from a different viewpoint is given by A. J. Maria and R. S. Martin, *Representation of positive harmonic functions*, Duke Mathematical Journal, vol. 2 (1936), pp. 517–529. A significant extension of the last results has recently been given by J. W. Green, *Harmonic functions in domains with multiple boundary points*, American Journal of Mathematics, vol. 61 (1939), pp. 609–632.

(2) It is scarcely necessary to mention that the present use of the term *minimal* bears no direct relationship to its use in connection with the problem of Plateau.

(3) See G. Bouligand, *Fonctions Harmoniques, Principes de Dirichlet et de Picard*, Mémorial des Sciences Mathématiques, no. 11, Paris, 1926; also, *Étude des singularités de certains champs scalaires*, Annales de l'École Normale Supérieure, (3), vol. 48 (1931), pp. 95–152.

to a fairly general form of the integral formula has been discussed by Maria and the author[4]. The general notion of a minimal function arises naturally when one considers from a more or less algebraic standpoint the *way* in which the integral formula represents the positive harmonic functions.

In space for a unit sphere with center at the point O, the formula in question is[5]

$$u(P) = \int_{\overline{OS}=1} F(S, P)d\mu(e_S) \qquad (\overline{OP} < 1),$$

where $F(S, P) = (1-\overline{OP}^2)/\overline{SP}^3$, and where $\mu(e)$ is a finite, non-negative, completely additive function of Borel sets (mass distribution function) on the surface of the sphere. The function $u(P)$ defined by this formula is always positive (non-negative) harmonic, and every such harmonic function is represented in this form by exactly one $\mu(e)$.

So far as the form of this representation is concerned, the two important features present are these: (1) A fixed family of functions, or *basis*, in terms of which the positive harmonic functions of the domain (sphere) are represented. This, of course, is the family of functions $F(S, P)$, where S plays the part of a parameter or index. (2) A linear process, or rather, a positively homogeneous and additive family of such processes analogous to and having as instance the formation of finite linear combinations with positive coefficients of functions from the basis. This is realized as the process of integration with respect to the mass distribution. Because of the properties of the Stieltjes integral with respect to a "point mass" distribution, this family of linear processes must contain all "unit combinations"—that is, all combinations of the form of a positive multiple of a single element of the basis—and the unit combinations, and only those, must be non-expressible as the sum of two linearly independent combinations[6]. In terms of these notions, there is a one-to-one representation of the positive harmonic functions of the domain as (generalized) positive linear combinations of functions from the basis.

Now the point to this formulation is that within the limitations described there is essentially only one basis that can be used in a representation of this type, namely, a suitably normalized family of minimal harmonic functions. More precisely, it can be shown that all functions of the basis must be positive harmonic and minimal in the domain and that exactly one positive multiple of each minimal function in the domain must occur in the basis. This is what the basis must be like, but it is, of course, not evident *a priori* that in a given

(4) Maria and Martin, loc. cit.

(5) See H. E. Bray and G. C. Evans, *A class of functions harmonic within the sphere*, American Journal Mathematics, vol. 49 (1927), pp. 153–180.

(6) Without this restriction it would seem that formal resemblance to the Stieltjes integral is for the most part destroyed. The point need not be pressed, since it is relevant only to the motivation of the present developments and not to the developments themselves.

domain minimal functions will exist, or, if they do, will exist in sufficient number to form the basis for such a representation. This leads to

PROBLEM A. *In a given domain, is the class of minimal functions sufficiently wide that, with a suitable normalization and a suitable definition of the linear process involved, it contains a basis for the positive harmonic functions of the domain?*

A central result of the present article is to give a general affirmative answer to this question; that is, to show that the answer is in the affirmative for an arbitrary domain. It is also shown that the linear process can always be realized by an integral of the Stieltjes type, and further, that every positive harmonic function of the domain is the limit of functions which are finite linear combinations with positive coefficients of minimal functions. Thus, in so far as the properties outlined above are considered indispensable to what ought to be considered a significant extension of the Poisson-Stieltjes integral formula, the present analysis, in view of its general applicability, serves as a background against which to examine critically the possibility of obtaining extensions which preserve other important features of the integral formula.

The arguments employed do not involve in an essential manner the dimensionality of the Euclidean space in which the domain is supposed to lie. Thus though the results here are explicitly for the three-dimensional case, only obvious modifications would be necessary for the others. For unbounded domains in the plane, the exceptional behavior of the logarithmic potential at infinity necessitates minor changes in some of the statements.

We sketch briefly the scheme of the argument. Returning for a moment to the Poisson-Stieltjes integral formula, we recall that the function $F(S, P)$ occurring there is actually the normal derivative at the point S of the Green's function $G(M, P)$ for the sphere. This normal derivative, not only in the case of the sphere but also in the case of any domain with a sufficiently smooth boundary, is equal (neglecting a positive factor independent of P) to the limit, as M approaches S, of a quotient of the form $G(M, P)/G(M, P_0)$, in which P_0 has been chosen as some fixed point of the domain. Approach here is not restricted to be along the normal at S; in fact, the quotient may have a well defined limit for all modes of approach to a boundary point even though at that point there is no normal. In the case, however, of a sufficiently irregular domain, there will be boundary points at which the limit of the quotient is not determinate. This suggests the introduction of ideal boundary elements. Speaking roughly, we identify an ideal boundary element with the totality of modes of approach to the boundary for which the quotient has a specified limit. This procedure is carried out in §2. With these ideal elements adjoined to the domain, we are able to obtain a convenient limiting form of the Riesz representation of superharmonic functions and, through it, an integral representation of positive harmonic functions bearing certain features of analogy

with the Poisson-Stieltjes formula (§3). The feature which this representation lacks is uniqueness; there may be more than one distribution representing a given harmonic function. The failure of the uniqueness is shown in §4 to be connected with the presence of non-minimal functions among the quotient limits. In the same section minimal functions are characterized, and it is shown that among all representations of a given harmonic function there is always exactly one (called canonical) which involves only minimal quotient limits. In this sense we recover the uniqueness of the representation and complete the answer to Problem A. In the concluding §5 certain applications and examples are treated.

1. **Auxilliary results on superharmonic functions.** We begin by recalling a number of results concerning the solution of the generalized Dirichlet problem(7). Let T be an open set in three-dimensional space; let t be its boundary. If P is a point of T, we denote by $m_T(e, P) = m(e, P)$ the mass distribution function resulting from sweeping a unit mass located at P out of T. P being fixed, $m(e, P)$ is a finite, non-negative, and completely additive function of sets e measurable Borel. The total mass is located upon t, more precisely upon the boundary of that component (maximal open connected subset) of T in which P lies, and is, in case this component is bounded, equal to unity; if P lies in an unbounded component of T, the total mass may be less. For fixed e, $m(e, P)$ is a non-negative harmonic function of P in T.

Let $\phi(Q)$ be a function defined for Q in t, measurable Borel, and summable over t with respect to $m(e, P)$ for each(8) P in T. The integral

$$u(P) = \int_t \phi(Q) dm(e_Q, P), \tag{1.1}$$

taken in the sense of Radon-Stieltjes-Lebesgue, defines a harmonic $u(P)$ in T. We speak of this $u(P)$ as *determined* in T by the boundary function $\phi(Q)$. In particular, if $\phi(Q)$ is continuous and, in case t is unbounded, approaches zero at infinity, the function $u(P)$ is identical in each component of T with the harmonic function determined there by the sequence solution, in the sense of Wiener, of the generalized Dirichlet problem for the boundary values $\phi(Q)$. Also under these circumstances $u(P)$ takes on continuously the boundary

(7) A rather complete bibliography of work relevant to this problem, together with an expository account, will be found in G. C. Evans, *Dirichlet problems*, American Mathematical Society Semicentennial Publications, vol. 2, New York, 1939, pp. 185–226. We shall cite: N. Wiener, *Certain notions in potential theory*, Journal of Mathematics and Physics (M.I.T.), vol. 3 (1924), pp. 127–146; C. de la Vallée Poussin, *Extension de la méthode du balayage de Poincaré et problème de Dirichlet*, Annales de l'Institut H. Poincaré, vol. 2 (1932), pp. 169–232; G. C. Evans, *Potentials of positive mass, parts* I *and* II, these Transactions, vol. 37 (1935), pp. 226–257, and vol. 38 (1935), pp. 201–236; O. Frostman, *Potentiel d'Équilibre et Capacité des Ensembles, avec Quelques Applications à la Théorie des Fonctions*, Thesis, Lund, 1935.

(8) Summability of the function for some one point P in each component of T is, of course, sufficient.

values $\phi(Q)$ except possibly at the irregular points of t, and approaches zero at infinity. The irregular points of t form at most a set of capacity 0[9].

Now let D be an arbitrary domain (open connected set) with boundary d. We shall assume in the sequel that D is fixed. Certain results concerning the positive superharmonic functions in D will be useful in the developments that follow.

THEOREM I. *Let $u(P)$ be positive and superharmonic in D. Let σ be a subset of D relatively closed in D. There exists uniquely a function $u_\sigma^*(P)$ defined in D such that:*

(a) *$u_\sigma^*(P)$ is superharmonic in D.*

(b) *$u_\sigma^*(P)=u(P)$ at all points of σ except possibly for those belonging to a subset of zero capacity.*

(c) *In $D-\sigma$, $u_\sigma^*(P)$ is identical with the function harmonic in $D-\sigma$ determined by the boundary function $\phi(Q)=\phi(\sigma, u; Q)$, where*

$$\phi(\sigma, u; Q) = \begin{cases} u(Q), & \text{when } Q \text{ is a boundary point of } D-\sigma \text{ lying in } D,\\ 0, & \text{when } Q \text{ is a boundary point of } D-\sigma \text{ in } d.\end{cases}$$

It is convenient to introduce the following notation: if $f(P)$ is defined in D, non-negative and Borel measurable there, we denote by $f(\sigma; P)$ the function defined as $f(P)$ for P in σ, and for P in $D-\sigma$ as the value at P of the harmonic function determined in $D-\sigma$ by the boundary function $\phi(\sigma, f; Q)$. It follows from this definition and the properties of an integral, in particular the integral of (1.1), that[10] if $f_n(P) \uparrow f(P)$, or $f_n(P)\to f(P)$, then correspondingly we have $f_n(\sigma; P) \uparrow f(\sigma; P)$, or $f_n(\sigma; P)\to f(\sigma; P)$, provided only in the latter case that there is a summable majorant for all the boundary functions involved.

Assume for the moment that the theorem is true. Consider the function $u(\sigma; P)$. This function and $u_\sigma^*(P)$, assumed to exist, agree except possibly on a subset of σ of zero capacity, *a fortiori* almost everywhere in D[11]. It follows that for P in any bounded subdomain completely interior to (contained with its boundary in) D, and for all sufficiently large n, $A_n u(\sigma; P)=A_n u_\sigma^*(P)$, where for an integrable $f(P)$ we denote by $A_n f(P)$ its integral mean (volume average) over a sphere of center P and radius $1/n$. Since $u_\sigma^*(P)$ is superharmonic[12], $A_n u_\sigma^*(P) \uparrow u_\sigma^*(P)$; hence $A_n u(\sigma; P) \uparrow u_\sigma^*(P)$. Thus for a proof

(9) Capacity is used here in the sense of de la Vallée Poussin. In this sense, a set is of zero capacity if every distribution of positive mass which is positive on the set generates an unbounded potential.

(10) The symbols $\uparrow$, $\downarrow$, and $\to$ will denote respectively increasing, decreasing, and unspecified convergence of sequences of numbers, functions, or point-sets. Context will determine the sense.

(11) Sets of zero capacity are necessarily of zero spatial measure.

(12) For a connected exposition of various known properties of super- (sub-) harmonic functions the reader is referred to the tract by T. Radó, *Subharmonic Functions*, Ergebnisse der Mathematik und ihrer Grenzgebiete, vol. 5, no. 1, Berlin, 1937.

of the theorem we are led to investigate the existence and properties of the limit of the sequence $\{A_n u(\sigma; P)\}$. Incidentally, the argument just given proves that if $u_\sigma^*(P)$ exists it is unique.

Suppose now that we have shown that (i) $u(\sigma; P)$ is lower semicontinuous at all points of D except possibly for a subset of σ of capacity zero, and that (ii) $u(\sigma; P)$ dominates its integral mean over any sphere of center P contained with its boundary in D. If D_1 is a bounded domain completely interior to D, there follows from (ii) for all P in D_1 and all sufficiently large n, $u(\sigma; P) \geqq A_n u(\sigma; P)$. For these n and for all P in a second domain D_2 completely interior to D_1, we then have for sufficiently large m, $A_m u(\sigma; P) \geqq A_m A_n u(\sigma; P) = A_n A_m u(\sigma; P)$. Thus $A_m u(\sigma; P)$, since continuous, is superharmonic in D_2. From this it follows, in particular, that if $n < p$, then $A_m A_n u(\sigma; P) = A_n A_m u(\sigma; P) \leqq A_p A_m u(\sigma; P) = A_m A_p u(\sigma; P)$. On making $m \to \infty$ in the first and last members of this relation and using the continuity of the average functions, we obtain $A_n u(\sigma; P) \leqq A_p u(\sigma; P)$ $(n < p)$. Since D_1 and D_2 are at our disposal, we have in D, $A_n u(\sigma; P) \uparrow u'(P) \leqq u(\sigma; P)$, where $u'(P)$ is superharmonic and where it is understood that approach of the prescribed type holds from some n on in any bounded domain completely interior to D. Now since the lower limit of the sequence of its average functions clearly dominates a function in any of its points of lower semicontinuity, it follows from this last relation and (i) that $u(\sigma; P)$ and $u'(P)$ are identical except possibly on a subset of σ of zero capacity. Thus $u'(P)$ satisfies the requirements for $u_\sigma^*(P)$. To prove the theorem it therefore suffices to demonstrate (i) and (ii).

Now let $u(P)$ and σ be as in the statement of the theorem. Let $f(P)$ be a function which is non-negative, continuous in $D+d$, zero on d, nowhere in D greater than $u(P)$, and which approaches zero at infinity if D is unbounded. Let Σ' be a bounded open subset of $\Sigma = D - \sigma$ completely interior to D and having only regular boundary points. Write $\sigma' = D - \Sigma'$, and consider $f(\sigma'; P)$. From the definition of a superharmonic function follows $f(\sigma'; P) \leqq u(P)$. If Σ' runs through an increasing sequence whose sum is Σ, then $f(\sigma'; P) \to f(\sigma; P)$. This is Wiener's result mentioned above. Thus we have $f(\sigma; P) \leqq u(P)$. Now $u(P)$, being non-negative and lower semicontinuous in D, may (e.g. by defining it as zero on d) be extended so as to have these properties on the closed set $D+d$. It follows from this that $u(P)$ can be approximated in D by an increasing sequence of functions satisfying the conditions imposed upon $f(P)$ above. Allowing $f(P)$ to run through such a sequence, we obtain $f(\sigma; P) \uparrow u(\sigma; P) \leqq u(P)$. At all points of D except possibly those irregular boundary points of Σ which are in D—and these form at most a subset of σ of zero capacity—the functions $f(\sigma; P)$ are continuous. Hence, with exactly the same possible exceptions, $u(\sigma; P)$ is lower semicontinuous in D. This establishes (i), and proves, incidentally, that $u(\sigma; P) \leqq u(P)$.

To prove (ii), it is convenient to write $v(P) = u(\sigma; P)$. Let Σ'' be any

bounded open set completely interior to D having only regular boundary points; write $\sigma''=D-\Sigma''$. We prove that $v(P)\geqq v(\sigma''; P)$. Assume that $f(P)$ is continuous in D, non-negative, and not greater than $v(P)$. From the definitions and $v(P)\leqq u(P)$, it follows that the difference $w(P)=v(P)-f(\sigma''; P)$ is non-negative in $\sigma+\sigma''$. In the remaining points of D, those of the open set $\Sigma\cdot\Sigma''$, the function $w(P)$ is harmonic and bounded from below. Since a boundary point of $\Sigma\cdot\Sigma''$ is in $\sigma+\sigma''$, $w(P)$ has a non-negative lower limit in any such boundary point which is also a point of lower semicontinuity of $v(P)$. As this means all boundary points except possibly those of a set of zero capacity, $w(P)$ is also non-negative in $\Sigma\cdot\Sigma''$(13); that is, $v(P)\geqq f(\sigma''; P)$. Now $v(P)$, though not necessarily lower semicontinuous, could be made so by modifying it on a set of zero capacity(14). Thus we can find an increasing sequence of continuous functions approaching $v(P)$ except on a set of zero capacity. On allowing $f(P)$ to run through such a sequence and observing that sets of zero capacity are null sets with respect to the swept-out mass in (1.1), we obtain for P in Σ'', $f(\sigma''; P)\uparrow v(\sigma''; P)$. This proves that $v(P)\geqq v(\sigma''; P)$, since the equality holds by definition in σ''.

Now let P_1 be any point of D; let S_ρ be the open sphere of center P_1 and radius ρ, where ρ is less than distance (P_1, d). Denote by s_ρ the boundary of S_ρ. If we take Σ'' above as S_ρ, the function $v(\sigma''; P)=v(D-S_\rho; P)$ is given for P in S_ρ by Poisson's integral with the boundary function $v(Q)$ on s_ρ. In particular, for $P=P_1$

$$v(P_1) \geqq v(D - S_\rho; P_1) = \frac{1}{4\pi\rho^2}\int_{s_\rho} v(Q)dQ,$$

where integration is with respect to area. On multiplying the first and last members of this relation by $4\pi\rho^2$ and integrating with respect to ρ between the limits 0 and r, where r is less than distance (P_1, d), we obtain

$$\tfrac{4}{3}\pi r^3\cdot v(P_1) \geqq \int_{S_r} v(P)dP,$$

integration now being with respect to volume. This establishes (ii), and completes the proof of Theorem I.

In the next theorem are listed a number of useful elementary properties of the function $u_\sigma^*(P)$ defined above. In the statement of the theorem, $u(P)$,

(13) If not, then for some negative number the set of boundary points at which the lower limit of the function does not exceed this number is of positive capacity. Cf. O. D. Kellogg, *Foundations of Potential Theory*, Berlin, 1929, p. 335. It may be noted that Kellogg uses the term *capacity* in the sense of Wiener, but for closed sets this coincides with the present usage.

(14) For example, this could be done by redefining the function as its lower limit at those points where it fails to be lower semicontinuous. These points form at most a set of zero capacity. Since the lower limit is not decreased in any point by this process, the modified function must be lower semicontinuous.

$v(P)$, etc. will be understood to denote non-negative superharmonic functions in D; σ, τ, etc., will be subsets relatively closed in D.

THEOREM II.

(a) $u(P) \geqq u_\sigma^*(P) \geqq 0$, *for all P in D.*

(b) *If $u(P) \geqq v(P)$ at all points of σ except for a subset of zero capacity, then $u_\sigma^*(P) \geqq v_\sigma^*(P)$ for all P in D.*

(c) $(u+v)_\sigma^*(P) = u_\sigma^*(P) + v_\sigma^*(P)$.

(d) $(c \cdot u)_\sigma^*(P) = c \cdot u_\sigma^*(P)$, *$c$ being a non-negative number.*

(e) *If $u_n(P) \to u(P)$ at all points of σ except for a subset of zero capacity, and if there exists a majorant $U(P)$ to the $u_n(P)$, where $U(P)$ is superharmonic in D, then $(u_n)_\sigma^*(P) \to u_\sigma^*(P)$ in D, except for a subset of σ of zero capacity. Lacking the majorant $U(P)$, we may still assert that* $\liminf_{n\to\infty} (u_n)_\sigma^*(P) \geqq u_\sigma^*(P)$ *at all points of $D-\sigma$.*

(f) *If $\sigma \subseteqq \tau$, then $(u_\tau^*)_\sigma^*(P) = (u_\sigma^*)_\tau^*(P) = u_\sigma^*(P)$.*

(g) *If $\sigma \subseteqq \tau$, then $u_\sigma^*(P) \leqq u_\tau^*(P)$. More generally, if $\sigma_n \uparrow \sigma$, then $u_{\sigma_n}^*(P) \uparrow u_\sigma^*(P)$.*

(h) $u_{\sigma+\tau}^*(P) \leqq u_\sigma^*(P) + u_\tau^*(P)$.

It is pertinent to make the obvious remark that if $u(P)$ and $v(P)$ are superharmonic, and $u(P) \geqq v(P)$ at almost all points of D, then consideration of volume averages extends the inequality at once to all points of D. Thus (a) holds, since it was established almost everywhere in the proof of Theorem I. The statements (b), (c), and (d) are easy consequences of the definitions and the remark just made.

For (e), a point P of σ where $(u_n)_\sigma^*(P) \to u_\sigma^*(P)$ fails must be a point where $u_n(P) \to u(P)$ fails, where $u_\sigma^*(P) \neq u(P)$, or where, for some n, $(u_n)_\sigma^*(P) \neq u_n(P)$. All such points form at most a subset of σ of zero capacity. In $D-\sigma$, in case the majorant $U(P)$ exists, the boundary functions $\phi(\sigma, u_n; Q)$ are dominated by the summable function $\phi(\sigma, U; Q)$, and the result is a consequence of Lebesgue's convergence theorem. If $U(P)$ fails to exist, we may still apply a well known lemma of Fatou.

Turning to (*f*), we observe that when $\sigma \subseteqq \tau$, $u_\tau^*(P)$ and $u(P)$ agree at all points of σ except those of a set of zero capacity. From (b) then follows $(u_\tau^*)_\sigma^*(P) = u_\sigma^*(P)$. Using this result in conjunction with (a) and (b), we obtain

$$(u_\sigma^*)_\tau^*(P) \geqq ((u_\sigma^*)_\tau^*)_\sigma^*(P) = (u_\sigma^*)_\sigma^*(P) = u_\sigma^*(P) \geqq (u_\sigma^*)_\tau^*(P),$$

which proves the other half of (f).

The first statement of (g) follows from (f) and (a). For the more general statement, write $v_n(P) = u_{\sigma_n}^*(P)$. Then $v_n(P) \uparrow v(P)$, where $v(P)$ is superharmonic and not greater than $u(P)$. In any point of σ where, for some n, $v_n(P) = u(P)$ holds, thus in all points of σ except for a set of zero capacity, $v(P) = u(P)$. From this and (e) it follows that $(v_n)_\sigma^*(P) \to u_\sigma^*(P)$ in D except

for a subset of σ of capacity zero. From (f), however, $(v_n)_\sigma^*(P) = v_n(P)$. This implies (with the aid of the remark at the outset of the proof) that $u_\sigma^*(P) = v(P)$ in D. Thus we have $v_n(P) \uparrow u_\sigma^*(P)$.

For (h), write $v(P) = u_\sigma^*(P) + u_\tau^*(P)$. It is clear from the non-negative character of the functions and from the definitions that $u(P) \leqq v(P)$ at all points of $\sigma + \tau$ except for a set of zero capacity. Thus using (b), (c), and (f) in succession, we have

$$\overset{*}{u}_{(\sigma+\tau)}(P) \leqq \overset{*}{v}_{(\sigma+\tau)}(P) = (\overset{*}{u_\sigma})^*_{(\sigma+\tau)}(P) + (\overset{*}{u_\tau})^*_{(\sigma+\tau)}(P)$$

$$= \overset{*}{u}_\sigma(P) + \overset{*}{u}_\tau(P).$$

We conclude the present section by obtaining for a special case the Riesz representation of the function $u_\sigma^*(P)$. It may be recalled that the generalized Green's function $G(M, P)$ for D is defined as $1/\overline{MP}$ minus the harmonic function $v_M(P)$ determined in D by the boundary function $\phi(Q) = 1/\overline{MQ}$. $G(M, P)$, thus defined for M and P in D, is non-negative and symmetric in its arguments. For fixed M it is harmonic in P except at $P = M$ and approaches zero at every regular point of d; it also approaches zero at infinity if D is unbounded.

THEOREM III. *If $u(P)$ is non-negative, superharmonic, and continuous in D, and if σ is a bounded closed subset of D, then*(15)

$$u_\sigma^*(P) = \int_\sigma G(M, P) d\nu_\sigma(e_M), \tag{1.2}$$

where $\nu_\sigma(e)$ is a finite, non-negative and completely additive function of Borel sets having its total mass in σ.

It is convenient to take Riesz's result in a form due to Frostman(16), who proved that the functional

$$J(\nu) = \frac{1}{2}\int_\sigma \int_\sigma G(M, P) d\nu(e_M) d\nu(e_P) - \int_\sigma u(P) d\nu(e_P),$$

under the same hypotheses upon $u(P)$ and σ as in the statement of the present theorem, is minimized by a unique $\nu(e) = \nu_\sigma(e)$ among all non-negative mass distributions $\nu(e)$ whose total mass is in σ. The function $v(P)$ given by

$$v(P) = \int_\sigma G(M, P) d\nu_\sigma(e_M)$$

(15) The hypotheses here are obviously unnecessarily restrictive, but the result in its present form is adequate for our purposes.

(16) O. Frostman, *La méthode de variation de Gauss et les fonctions sousharmoniques*, Acta Litterarum ac Scientiarum Regiae Universitatis Hungaricae Francisco-Iosephinae, Szeged, vol. 8 (1936–1937), pp. 116–126.

is equal to $u(P)$ at all points of σ except possibly for those of a subset of capacity zero, and nowhere in D exceeds $u(P)$. The proof of the present theorem reduces to showing that $v(P)=u_\sigma^*(P)$.

The function $v(P)$ is superharmonic in D, harmonic in $D-\sigma$ and approaches zero at every regular point of d and also at infinity in case D is unbounded. From the inequality $v(P)\leqq u(P)$, the continuity of $u(P)$, and the lower semicontinuity of $v(P)$, follows the continuity of $v(P)$ in any point where $v(P)=u(P)$. Thus in $D-\sigma$ the function $v(P)$ approaches boundary values $\phi(\sigma, u; Q)$ at all boundary points of $D-\sigma$ except possibly for a set of zero capacity. The function $u_\sigma^*(P)$ satisfies exactly the same boundary conditions in $D-\sigma$, and, since both functions are bounded, they are identical there. In σ the functions differ on at most a set of zero capacity; hence they are identical in D.

2. **The ideal boundary elements and the metric ρ.** In this section we define for D a set of ideal boundary elements and derive certain properties of the domain with these elements adjoined.

Let P_0 be a point of D chosen arbitrarily but fixed for the ensuing discussion([17]). We denote by $G(M, P)$ the generalized Green's function for D, and define for M and P in D

$$K(M,P) = \begin{cases} G(M, P)/G(M, P_0) & (M \neq P_0), \\ 0 & (M = P_0;\ P \neq P_0), \\ 1 & (M = P = P_0). \end{cases}$$

The function $K(M, P)$ is for fixed M a non-negative harmonic function of P except at $P=M$; its value at $P=P_0$ is 1. For fixed P, $K(M, P)$ is continuous as a function of M except at $M=P$.

Consider now a sequence $\{M_n\}$ of points of D having no point of accumulation in D. In any bounded closed subdomain of D, the functions $K(M_n, P)$ form, from some n on, a bounded sequence of harmonic functions of P—thus a normal family([18]). A subsequence of these functions, therefore, is convergent in D to a positive harmonic function. A sequence $\{M_n\}$ of points of D having no accumulation point in D, for which the corresponding $K(M_n, P)$ have the property of the subsequence just mentioned—that is, converge to a harmonic function in D—will be called *fundamental*. We have just seen that any se-

([17]) As will be seen presently, the particular choice of this point makes no essential difference in the structure of the ideal boundary or in the theory that follows. In fact, the only effect of a change in the choice is to multiply the functions $K(M, P)$ by a factor $1/K(M, P_0')$, where P_0' is the new normalizing point.

([18]) The normalization at P_0 entails the boundedness both from above and positively from below of the function $K(M, P)$ for P in a bounded domain containing P_0 uniformly in M when M is kept more than a fixed positive distance from this domain. Kellogg, loc. cit., pp. 263–265. The normal family property is a well known consequence of Harnack's inequality and the theorem of Ascoli.

quence of points of D without accumulation point in D has a fundamental subsequence. Two fundamental sequences are called *equivalent* if their corresponding $K(M, P)$'s have the same limit. This has the usual properties of an equivalence relation.

DEFINITION 1. *The class of all fundamental sequences equivalent to a given one determines (or, simply, is) an ideal boundary element of D. The set of all ideal boundary elements of D will be denoted by Δ, and the set $D+\Delta$, by $\mathfrak{D}$.*

The domain of definition of $K(M, P)$ may now be extended by writing

$$K(M, P) = \lim_{n\to\infty} K(M_n, P) \qquad (M \text{ in } \Delta;\ P \text{ in } D),$$

where $\{M_n\}$ is any fundamental sequence determining M. For M in Δ, $K(M, P)$ is thus a positive harmonic function of P in D having the value 1 for $P=P_0$.

Evidently the function $K(M, P)$ is characteristic of the point M in the sense that the identity of two points of $\mathfrak{D}$ is equivalent to the equality of their corresponding $K(M, P)$'s as functions of P. Thus, in view of the possible application of the normal family theory, it is to be expected that $\mathfrak{D}$ can be given a topology with respect to which it is compact and with respect to which $K(M, P)$ as a function of M possesses certain continuity properties. This may be shown more explicitly by introducing a metric. Actually, the precise analytic form of the metric we choose to introduce is not of great importance in the present developments but has some technical advantages.

DEFINITION 2. *Select a fixed sphere Σ completely interior to D having, say, P_0 for center. For M and M' in $\mathfrak{D}$ we define*[10]

$$\rho(M, M') = \int_\Sigma \frac{|K(M, P) - K(M', P)|}{1 + |K(M, P) - K(M', P)|}\, dP, \tag{2.1}$$

in which integration is with respect to volume and in which the integrand is defined conventionally if $P = M$ or M'.

THEOREM I. *The function $\rho(M, M')$ is a metric in $\mathfrak{D}$. With respect to it $\mathfrak{D}$ is complete and compact, D is open, and Δ is the boundary of D. The relative topology in D arising from the metric is equivalent to the original topology there.*

That $\rho(M, M')$ is finite, non-negative, symmetric, and that it satisfies the triangular inequality and vanishes if $M=M'$ is clear from (2.1). If $\rho(M, M')=0$, the integrand in (2.1), since non-negative, must vanish in every point of Σ at which it is continuous. Hence, $K(M, P)=K(M', P)$ for all P in Σ except possibly $P=M$ or $P=M'$. Harmonic continuation extends

(10) Taking the integrand in this form is, of course, a purely technical device for obtaining a bounded integrand. Cf. S. Banach, *Théorie des Opérations Linéaires*, Warsaw, 1932, pp. 9–10.

the equality to all P in D. Thus $\rho(M, M')=0$ implies $M=M'$. This establishes the first statement of the theorem.

The remainder of the proof will be carried out in a number of brief steps. A sequence $\{M_n\}$ of points of $\mathcal{D}$ convergent to a point M in the sense of the metric ρ will be called *ρ-convergent* to M.

(i) *A sequence $\{M_n\}$ of points of D convergent to a point M of D is ρ-convergent to M.* For, if in (2.1) M is taken as the present M, and M' is replaced successively by $M_1, M_2, \cdots$, the integrands are bounded and converge to zero except possibly at a single point.

(ii) *A fundamental sequence $\{M_n\}$ determining a point M of Δ is ρ-convergent to M.* This follows from the same argument as in (i) and the definition of a fundamental sequence.

(iii) *A sequence $\{M_n\}$ of points of Δ has a subsequence ρ-convergent to a point M of Δ.* Let $F_1, F_2, \cdots$ be an increasing sequence of bounded closed sets whose sum is D. Consider a fixed M_n. A fundamental sequence determining M_n has at most a finite number of points in F_n. We can, by (ii), select from such a fundamental sequence a point which is not in F_n and whose ρ-distance from M_n does not exceed $1/n$. Call this point M_n'. We thus obtain a sequence $\{M_n'\}$ of points of D which, since M_n' is not in F_n, can have no accumulation point in D, and for which $\rho(M_n, M_n') \leqq 1/n$. A subsequence of $\{M_n'\}$ is fundamental, and determines an M in Δ. Application of the last inequality, the triangular inequality, and (ii) shows that the corresponding subsequence of $\{M_n\}$ is ρ-convergent to M.

(iv) *Any sequence $\{M_n\}$ of points of $\mathcal{D}$ has a subsequence ρ-convergent to a point M of $\mathcal{D}$.* If an infinity of points of $\{M_n\}$ are in D, and these have an accumulation point M there, a subsequence consists of points of D and converges to M. We then apply (i). If an infinity of points of $\{M_n\}$ are in D but there is no point of accumulation there, a subsequence is fundamental, and we apply (ii). This leaves only the possibility that an infinity of the M_n are in Δ; (iii) applies here.

(v) *A sequence $\{M_n\}$ of points of D which is ρ-convergent to a point M of D is convergent in the ordinary sense to M.* For, if the sequence $\{M_n\}$ did not converge to M, it would either have a subsequence convergent to an $M' \neq M$ in D, or would have a fundamental subsequence. In the first case (i), and in the second (ii), would imply a contradiction.

The various statements of the theorem now follow at once. (iv), as the statement of self-compactness, implies completeness and compactness. That Δ is ρ-closed follows from (iii). D, as the complement of Δ in $\mathcal{D}$, is ρ-open. From (ii) and the fact that a fundamental sequence consists only of points of D, it follows that every ρ-neighborhood of a point of Δ contains points of D. The equivalence of the two topologies in D is a consequence of (i) and (v).

THEOREM II. *The function $K(M, P)$, for fixed P, is ρ-continuous as a func-*

tion of M in $\mathfrak{D}$, except at $M=P$. More generally, if F is a bounded closed subset of D and if G is a ρ-closed subset of $\mathfrak{D}-F$, then $K(M, P)$ is uniformly continuous in both arguments for M in G and P in F.

The first statement, when M is in D, is a consequence of the equivalence of continuity and ρ-continuity in D. Suppose that M is in Δ, and let $\{M_n\}$ be any sequence of points of $\mathfrak{D}$ ρ-convergent to M. For a subsequence $\{M_n'\}$ of $\{M_n\}$, we have $K(M_n', P)\to v(P)$, where $v(P)$ is harmonic in D. If in (2.1) M is taken as the present M and M' is replaced successively by $M_1', M_2', \cdots$, then the condition $\rho(M, M_n')\to 0$ implies that the integral in (2.1) with $K(M', P)$ replaced by $v(P)$ has the value zero. Thus $v(P)=K(M, P)$ for all P in Σ, and harmonic continuation extends this equality to all P in D. In other words, $\{M_n\}$ has a subsequence $\{M_n'\}$ such that $K(M_n', P)\to K(M, P)$. Since the same argument applies to any subsequence of $\{M_n\}$, it follows that $K(M_n, P)\to K(M, P)$.

For the second statement, consider the $K(M, P)$ with M in G as a family of harmonic functions of P in $D-G$. They form a family uniformly bounded near any point of $D-G$; thus, since harmonic, they are equicontinuous at any such point, in particular at any point of F. Hence, for M in G and P in F, $K(M, P)$ is continuous in M and continuous in P uniformly in M. Continuity in both arguments follows from this and the compactness of the ranges of M and P.

The notions of ρ-closed and ρ-open sets arising from the metric ρ in $\mathfrak{D}$ extend in a familiar manner to that of a *ρ-Borel set*. A system of sets in a space is customarily called a *Borel field* if it contains the empty set, contains with each of its sets the complement, and contains with each sequence of its sets the sum[20]. The system of ρ-Borel sets of $\mathfrak{D}$ is defined as the smallest Borel field consisting of sets in $\mathfrak{D}$ which contains all ρ-open sets.

THEOREM III. *The ρ-Borel subsets of D are identical with the subsets of D measurable Borel in the ordinary sense.*

Consider the system $\mathfrak{S}$ of sets of $\mathfrak{D}$ whose intersection with D is a Borel set. $\mathfrak{S}$ is clearly a Borel field. Since the intersection of a ρ-open set with D is ρ-open and, hence, by Theorem I, open, $\mathfrak{S}$ contains all ρ-open sets, and thus all ρ-Borel sets. It follows that every ρ-Borel subset of D is Borel.

Conversely, a similar consideration of the system $\mathfrak{S}'$ of all sets in space whose intersection with D is ρ-Borel shows that $\mathfrak{S}'$ contains all Borel sets in space. In particular, every Borel subset of D is ρ-Borel.

Relative to any Borel field there are definable the notions of a completely additive function of sets of the field, of measurability of a point function with respect to the field, and of an integral of a measurable point function with

---

(20) See, for example, S. Saks, *Theory of the Integral*, 2d revised edition, translated by L. C. Young, Warsaw, 1937, p. 7. Saks uses the term *additive class*.

respect to a completely additive function of sets[21]. In particular, we shall make use of these notions relative to the system of ρ-Borel sets. Since $\mathfrak{D}$ is ρ-compact, the theory of such an integral does not differ essentially from that a Radon-Stieltjes-Lebesgue integral over a bounded closed portion of Euclidean space.

We shall have occasion to use only those completely additive set functions which are finite and non-negative, and shall use the terms *mass distribution* or *mass function* only in reference to such. The notion of *weak convergence* of a sequence of mass distributions is of considerable importance, and may be recalled. If F is a closed and compact set in a metric space, a distribution $\mu(e)$ over F is called the *weak limit* of a sequence $\{\mu_n(e)\}$ of distributions over F, if the condition

$$\lim_{n\to\infty} \int_F f(S)d\mu_n(e_S) = \int_F f(S)d\mu(e_S) \tag{2.2}$$

holds for every function $f(S)$ which is continuous over F. The most important feature of weak convergence is the

SELECTION THEOREM[22]. *A sequence of distributions over a closed compact set F in which the total masses are uniformly bounded has a subsequence weakly convergent to a distribution over F.*

3. **The function $u_A(P)$ and the representation.** Throughout this section and the next we shall employ the metric ρ in $\mathfrak{D}$. Thus, the terms, *open*, *closed*, *distance*, etc., will be understood, in the absence of specific mention to the contrary, in the sense of this metric. If G is any set in $\mathfrak{D}$, we shall denote by $[G]$ the intersection of D with the ρ-closure of G. $[G]$ is a relatively closed subset of D in the sense of the ordinary topology of D.

---

(21) The integral most convenient for our present purposes is one of Stieltjes type; that is, one whose dependence upon a mass distribution is explicit. The theory in the extended form needed here may be found in Saks, op. cit.

(22) The definition of weak convergence given above is copied after that of J. Radon, *Theorie und Anwendungen der absolut additiven Mengenfunktionen*, Sitzungsberichte der Akademie der Wissenschaften, Vienna, 1913, p. 1337. All that is really essential for the extension of the result is the observation that, since we are in a metric space and F is compact, there is a countable neighborhood system covering F. It may be noted that the condition (2.2) is equivalent to three: (i) the $\mu_n(F)$ are bounded independently of n; (ii) $\liminf_{n\to\infty} \mu_n(e) \geqq \mu(e)$, for subsets e open in F; (iii) $\limsup_{n\to\infty} \mu_n(e) \leqq \mu(e)$, for closed subsets e of F. It is possible to imbed the countable neighborhood system of F in a system of sets $\mathfrak{P}$, also countable, containing with each of its sets the complement in F, and with each pair of its sets the sum and intersection. By the "diagonal process" we can determine a subsequence of the mass distributions having a limit $\lambda(p)$ for each set p of $\mathfrak{P}$. $\lambda(p)$ is additive in the restricted sense for sets in $\mathfrak{P}$, and can be used to generate a Carathéodory outer measure expressible in a familiar fashion in terms of coverings by sequences of sets from $\mathfrak{P}$. The measure function of this outer measure, when restricted to Borel sets, satisfies the requirements (ii), and (iii) for $\mu(e)$.

DEFINITION 1. *Let $u(P)$ be a non-negative harmonic function in D, and let A be a closed subset of Δ. The function $u_A(P)$ is defined as the greatest lower bound of $u^*_{[G]}(P)$ as G ranges over all open sets containing A.*

LEMMA 1. *Let $u(P)$ and A be as in the above definition. Let $G_1, G_2, \cdots$ be a descending sequence of open sets which contain A and whose closures have A as their intersection. Then $u^*_{[G_n]}(P) \downarrow u_A(P)$.*

Obviously, the sets $[G_n]$ form a descending sequence having a void intersection. The functions $u^*_{[G_n]}(P)$ form a descending sequence and, from some n on, are non-negative harmonic in any bounded closed subset of D. Thus, $u^*_{[G_n]}(P) \downarrow v(P)$, where $v(P)$ is non-negative harmonic in D. From Definition 1, we have $v(P) \geqq u_A(P)$. On the other hand, corresponding to any point P and any positive ϵ, there is an open set G containing A such that, for this particular P, we have $u^*_{[G]}(P) \leqq u_A(P)+\epsilon$. But, since G will contain all but a finite number of the sets G_n, we have $u^*_{[G]}(P) \geqq v(P)$. Combining these inequalities and using the arbitariness of ϵ, we obtain $v(P) = u_A(P)$.

We now derive a number of elementary properties of the function $u_A(P)$. In the statement of the following theorem, $u(P)$ and $v(P)$ will denote non-negative harmonic functions in D; A, B, etc. will be closed subsets of Δ.

THEOREM I. *The function $u_A(P)$ is non-negative harmonic in D. It has the following properties*(22):

(a) $u(P) \geqq u_A(P)$ *for all P in D.*

(b) *If $u(P) \geqq v(P)$ for all P in D, then $u_A(P) \geqq v_A(P)$.*

(c) $(u+v)_A(P) = u_A(P)+v_A(P)$.

(d) $(c\cdot u)_A(P) = c\cdot u_A(P)$, *where c is a non-negative number.*

(e) $u_\Delta(P) = u(P)$.

(f) *If $A \supseteqq B$, then $(U_B)_A(P) = u_B(P)$.*

(g) *If $A \supseteqq B$, then $u_A(P) \geqq u_B(P)$. More generally, if $A_n \downarrow A$, then $u_{A_n}(P) \downarrow u_A(P)$.*

(h) $u_{(A+B)}(P) \leqq u_A(P)+u_B(P)$.

That $u_A(P)$ is harmonic was shown incidentally in the proof of Lemma 1. Statements (a), (b), (c), and (d) are immediate consequences of Lemma 1 and the corresponding statements in Theorem II of §1.

To prove (e), suppose that G is an open set containing Δ. $\mathfrak{D}-G$ is a subset of D, is ρ-closed and, therefore, is closed in the ordinary sense. It is also bounded. If it were not, there could be selected from it a sequence of points having no point of accumulation in D; this sequence would contain a fundamental subsequence determining a point of Δ. This is impossible. From this it follows that $D-[G]$ is a bounded open set completely interior to D. In

(22) Several of the statements here are shown by Theorem III, §4, and its Corollary 3 to admit of considerable strengthening.

every boundary point of $D-[G]$, $u(P)$ is harmonic; thus it is determined in $D-[G]$ by its own boundary values. It follows that in $D-[G]$, $u^*_{[G]}(P)=u(P)$. (e) now follows if we let G close down upon Δ through a sequence of the type prescribed in the statement of Lemma 1.

To prove (f), let G and H be open sets containing A and B respectively, and assume that $G\supseteq H$. From (f) of Theorem II, §1, we have $(u^*_{[H]})^*_{[G]}(P)=u^*_{[H]}(P)$. If we keep G fixed and allow H to run through a descending sequence of sets closing down upon B as in Lemma 1, we obtain, with the aid of (e) of Theorem II, §1 (using $u(P)$ for a majorant), $(u_B)^*_{[G]}(P)=u_B(P)$ for all P in $D-[G]$. (f) now follows as the limiting form of this last relation when G is allowed to run through a descending sequence of sets closing down upon A.

The first statement of (g) follows from Definition 1 and the fact that any open set containing A also contains B. For the more general statement, the functions $u_{A_n}(P)$ form a descending sequence having a harmonic limit $v(P)$, which clearly must dominate $u_A(P)$. On the other hand, since any open set G containing A contains all but a finite number of the A_n, we have, for such G, $u^*_{[G]}(P)\geqq v(P)$. It follows that $u_A(P)\geqq v(P)$, and therefore that $u_A(P)=v(P)$.

To prove (h), let G and H be open sets containing A and B respectively. From (h) of Theorem II, §1, we have, writing $K=G+H$ and observing that $[K]=[G]+[H]$, $u^*_{[K]}(P)\leqq u^*_{[G]}(P)+u^*_{[H]}(P)$. If now G and H run simultaneously through descending sequences of sets closing down respectively upon A and B, K runs through a sequence closing down upon $A+B$, and (h) follows as the limiting form of the last inequality.

THEOREM II. *If $u(P)$ is non-negative harmonic in D, and A is a closed subset of Δ, then there exists a mass distribution $\mu_A(e)$ over A such that*

$$u_A(P) = \int_A K(M, P)d\mu_A(e_M) \tag{3.1}$$

for all P in D. The total mass, $\mu_A(A)$, is equal to the value of the function $u_A(P)$ at the point P_0.

It should be remarked that the uniqueness of the distribution $\mu_A(e)$ is *not* asserted in this statement. As we shall see later, there may actually be more than one distribution satisfying the requirements of the theorem.

Let G be an open set containing A and having the point P_0 as an exterior point. Denote by $\overline{G}$ the ρ-closure of G. Let σ be a closed subset of $[G]$. σ is then closed and bounded (cf. the proof of part (e) of Theorem I of the present section) in the ordinary topology of D. If we now transform the integral of (1.2), and use the fact that $G(M, P_0)$ as a function of M in σ is continuous and positive, we obtain

$$(3.2)\qquad \begin{aligned} u_\sigma^*(P) &= \int_\sigma G(M, P)d\nu_\sigma(e_M) \\ &= \int_\sigma K(M, P)\cdot G(M, P_0)d\nu_\sigma(e_M) = \int_\sigma K(M, P)d\mu_\sigma(e_M), \end{aligned}$$

in which

$$\mu_\sigma(e) = \int_e G(M, P_0)d\nu_\sigma(e_M).$$

Since $\mu_\sigma(e)$ is a Borel mass distribution vanishing outside σ, it may, by the result of Theorem III of §2, equally well be interpreted as a ρ-Borel mass distribution over σ. Since $\sigma \subset \overline{G}$, the result of (3.2) may be written

$$(3.3)\qquad u_\sigma^*(P) = \int_{\overline{G}} K(M, P)d\mu_\sigma(e_M).$$

The total mass of $\mu_\sigma(e)$, calculated by writing $P = P_0$ in (3.3) and recalling that $K(M, P_0) \equiv 1$, is $u_\sigma^*(P_0)$, and thus does not exceed $u(P_0)$.

If now σ is allowed to run through an ascending sequence of sets whose sum is $[G]$, then, since the $\mu_\sigma(e)$ have total masses not exceeding $u(P_0)$ and lying in the closed compact set $\overline{G}$, we have for some subsequence of $\{\mu_\sigma(e)\}$—and we may assume it already extracted—a weak limit distribution $\mu_{\overline{G}}(e)$ having its total mass (also not greater than $u(P_0)$) in $\overline{G}$. From Theorem II, (g), §1, we have $u_\sigma^*(P) \uparrow u_{[G]}^*(P)$. Using the weak convergence of the distributions and the continuity of $K(M, P)$ as a function of M in $\overline{G}$ when P is in $D - [G]$, we obtain as the limiting form of the equation (3.3)

$$(3.4)\qquad u_{[G]}^*(P) = \int_{\overline{G}} K(M, P)d\mu_{\overline{G}}(e_M) \qquad (P \text{ in } D - [G]).$$

Now allow G to run through a sequence closing down upon A in the manner prescribed in the statement of Lemma 1. From the result of Lemma 1, by an argument similar to that of the preceding paragraph coupled with the observation that a weak limiting distribution, $\mu_A(e)$, of the $\mu_{\overline{G}}(e)$ must have its total mass in every $\overline{G}$ and thus in A, we obtain (3.1) as the limiting form of (3.4). The last statement of the theorem is immediate if we put $P = P_0$ in (3.1).

We now have a representation theorem:

THEOREM III. *If $u(P)$ is non-negative harmonic in D, then there exists a distribution $\mu(e)$ over Δ such that*

$$(3.5)\qquad u(P) = \int_\Delta K(M, P)d\mu(e_M),$$

for all P in D. Conversely, for any distribution $\mu(e)$ over Δ, the integral in (3.5)

represents a non-negative harmonic function $u(P)$. The total mass, $\mu(\Delta)$, is equal to the value of $u(P)$ at the point P_0.

The first and last statements follow as corollaries of Theorem I, (e), and Theorem II of the present section if A is taken as Δ in the latter theorem.

For the converse statement, since $K(M, P)$ is continuous as a function of M in Δ, the integral in (3.5) can be approximated by means of Riemann sums. Indeed, in view of Theorem II, §2, the approximation is uniform for P in any bounded closed subset of D. The approximating sums are finite positive linear combinations of $K(M, P)$'s with M in Δ, and, as such, are non-negative harmonic. The result is now immediate.

4. **The minimal functions and the uniqueness problem.** As has been indicated before, the representation obtained in the preceding section fails to give a complete determination of the distribution in terms of which a specified harmonic function is represented. In this section we shall obtain a characterization of the minimal harmonic functions and establish the existence of a unique canonical representation in terms of these functions.

LEMMA 1. *Suppose that $u(P)$ is positive harmonic and minimal. Let A be any ρ-Borel subset of Δ. If now a relation of the form*

$$u(P) \geqq \int_A K(M, P)d\mu(e_M) > 0 \tag{4.1}$$

obtains for all P in D, then $u(P) = u(P_0)\cdot K(S, P)$, where S is some point in A.

$\mu(A)$ is positive, as is easily seen by setting $P = P_0$ in (4.1). A, therefore, has a closed subset A_1 for which $\mu(A_1)$ is positive. A_1, being compact, can be covered by a finite number of its closed subsets, all of them having diameter less than some selected positive number. At least one such subset has a positive μ mass. We select a particular such and call it A_2. By proceeding in this way inductively, it is possible to construct a descending sequence $A_1, A_2, \cdots$ of closed subsets of A whose diameters approach zero and each of which has a positive μ mass. Let S be the (unique) point common to all the A_n.

Now since $\mu(A_n)$ is positive, the integral in (4.1) extended over A_n instead of A represents a positive harmonic function dominated by the minimal function $u(P)$, and is thus equal to $c_n \cdot u(P)$, where c_n is positive. If we write $\mu_n(e) \equiv c_n^{-1}\cdot\mu(A_n\cdot e)$, there follows

$$u(P) = \int_{A_n} K(M, P)d\mu_n(e_M). \tag{4.2}$$

The total mass of the distribution $\mu_n(e)$ is $u(P_0)$ and is located upon A_n. Thus the $\mu_n(e)$ have as weak limit a *point mass* of amount $u(P_0)$ located at S. The relation $u(P) = u(P_0)\cdot K(S, P)$ now follows as the limiting form of (4.2).

COROLLARY 1. *Every minimal positive harmonic function in D is a positive multiple of some $K(S, P)$, where S is in Δ.*

This follows on taking $A=\Delta$ in the lemma and recalling the result of Theorem III of the preceding section.

COROLLARY 2. *If $K(S, P)$ is minimal and A is a closed subset of Δ such that $K_A(S, P)$ is positive, then S is in A.*

Since $K(S, P)\geqq K_A(S, P)$, Theorem II of the preceding section with A taken as the present A implies a relation of the form (4.1).

DEFINITION 1. *We define the function $\psi(S)$ for S in Δ as $K_{\{S\}}(S, P_0)$; that is, as the value at P_0 of $u_A(P)$, where $u(P)$ is $K(S, P)$ and A is the set consisting of the single point S.*

THEOREM I. *The function $\psi(S)$ has only the two possible values* 1 *and* 0. *The function $K(S, P)$ is minimal or not according as $\psi(S)$ is equal to* 1 *or* 0.

Theorem II of the preceding section with $A=\{S\}$ has the following consequence:

$$u_{\{S\}}(P) = \int_{\{S\}} K(M, P)d\mu_{\{S\}}(e_M) = u_{\{S\}}(P_0)\cdot K(S, P). \tag{4.3}$$

In particular, putting $u(P)=K(S, P)$, we get

$$K_{\{S\}}(S, P) = \psi(S)\cdot K(S, P).$$

Application to this of (f) and (d) Theorem I, §3, yields

$$K_{\{S\}}(S, P) = (K_{\{S\}})_{\{S\}}(S, P) = \psi(S)\cdot K_{\{S\}}(S, P).$$

If we put $P=P_0$, this yields $\psi(S)=[\psi(S)]^2$, which proves that $\psi(S)=1$ or 0.

Assume now that $\psi(S)=1$. We show that this implies that $K(S, P)$ is minimal. Let $u(P)$ be any positive harmonic function dominated by $K(S, P)$. Write $v(P)=K(S, P)-u(P)$. Then $v(P)$ is non-negative harmonic. Now

$$u(P) \geqq u_{\{S\}}(P), \qquad v(P) \geqq v_{\{S\}}(P), \tag{4.4}$$

and

$$\begin{aligned} K(S, P) &= u(P) + v(P) \geqq u_{\{S\}}(P) + v_{\{S\}}(P) \\ &= K_{\{S\}}(S, P) = \psi(S)\cdot K(S, P) = K(S, P). \end{aligned} \tag{4.5}$$

But clearly (4.5) can hold only if equality obtains in both of the relations (4.4). In particular, $u(P)=u_{\{S\}}(P)$; from (4.3) the latter function is a multiple of $K(S, P)$. Thus $K(S, P)$ is minimal.

Assume, conversely, that $K(S, P)$ is minimal. Suppose that A is a closed subset of Δ having S as an interior point relative to Δ; let B be the closure

of the complement of A in Δ. We now have $K_B(S, P)=0$; for, if not, Corollary 2 above would imply that S is in B. Thus,

$$K(S, P) = K_\Delta(S, P) = K_{(A+B)}(S, P)$$
$$\leqq K_A(S, P) + K_B(S, P) = K_A(S, P) \leqq K(S, P);$$

that is, $K_A(S, P)=K(S, P)$. If now A is allowed to run, for example, through the sequence $\{A_n\}$, where A_n is the set of points of Δ whose ρ-distance from S does not exceed $1/n$, then $A_n \downarrow \{S\}$, and we have, as the limiting form of the result just obtained, $K_{\{S\}}(S, P)=K(S, P)$. On writing $P=P_0$ in this, there follows $\psi(S)=1$.

Definition 2. *We shall denote by* Δ_0 *and* Δ_1 *respectively the sets of points of* Δ *for which* $\psi(S)$ *has the value* 0 *and the value* 1.

Theorem II. *The set* Δ_0 *is either void, or closed, or an* F_σ.

We introduce an auxiliary sequence of sets Γ_n, where $n=1, 2, \cdots$. The set Γ_n is defined as the set (possibly void) of all points S of Δ having the following property: If G is any open set in $\mathcal{D}$ containing S and having a ρ-diameter less than $1/n$, then $K^*_{[G]}(S, P_0)\leqq \frac{1}{2}$. It is clear, incidentally, from this definition that the Γ_n form an ascending sequence.

Since $K(S, P)$ is continuous as a function of S in Δ, it follows from the second statement in (e) of Theorem II, §1 that for any open set G the function $K^*_{[G]}(S, P_0)$ is lower semicontinuous as a function of S. In particular, if S_0 is a limit point of Γ_n (assumed non-void) and G is an open set of diameter less than $1/n$ containing S_0, then the function $K^*_{[G]}(S, P_0)$ has a value not exceeding $\frac{1}{2}$ at points S of a sequence approaching S_0. It follows that S_0 is in Γ_n; that is, Γ_n is closed.

Now let S be a point of Γ_n. Select an open set G (e.g., an open ρ-sphere) containing S and having a diameter less than $1/n$. There follows $\psi(S) =K_{\{S\}}(S, P_0)\leqq K^*_{[G]}(S, P_0)\leqq \frac{1}{2}<1$; that is, $\psi(S)=0$. This proves that $\Gamma_n \subseteqq \Delta_0$ $(n=1, 2, \cdots)$.

Assume, conversely, that S is in Δ_0. Denote by G_n the set of points of $\mathcal{D}$ whose distance from S is less than $1/n$. Then, since $\overline{G}_n \downarrow \{S\}$, we have $K^*_{[G_n]}(S, P_0) \downarrow K_{\{S\}}(S, P_0)=\psi(S)=0$. Hence, we may choose $n=m$ so great that $K^*_{[G_m]}(S, P_0)\leqq \frac{1}{2}$. Any open set G which contains S and has a diameter less than $1/m$ is contained in G_m. Since for any such G, $K^*_{[G]}(S, P_0) \leqq K^*_{[G_m]}(S, P_0)$, it follows that S is in Γ_m.

The result of the last two paragraphs is to the effect that Δ_0 is identical with the sum of the Γ_n, which were proved above to be closed or void[24].

[24] It seems a reasonable conjecture that the sets Γ_n are nowhere dense in Δ; thus that Δ_0, actually, is of the first category in Δ. An answer to this question would have interesting consequences.

The particular consequence of the theorem just proved, that Δ_0 is a ρ-Borel set, is needed in

DEFINITION 3. *A distribution $\mu(e)$ over Δ will be called canonical if $\mu(\Delta_0)=0$. A representation of the form given by Theorem* III, §3, *is a canonical representation if the distribution occurring in it is canonical.*

In a canonical distribution the total mass is carried on the set Δ_1. Thus a canonical representation is one which involves, in a sense, only minimal $K(S, P)$'s. It is our purpose to show that every non-negative harmonic function in D has exactly one canonical representation. Before proceeding to this result it is convenient to prove a number of lemmas.

LEMMA 2. *If Γ_n is one of the auxiliary sets introduced in the proof of Theorem* II, *then $u_{\Gamma_n}(P)=0$, for any function $u(P)$ positive harmonic in D.*

The set Γ_n, being closed and compact, may be covered by a finite number of its closed subsets each of diameter less than $1/n$. It is sufficient (Theorem I, (h), §3) to prove that $u_A(P)=0$ whenever A is such a subset of Γ_n. A being such a set, let G be an open set also of diameter less than $1/n$ containing A. From the defining property of Γ_n, we have $K^*_{[G]}(S, P_0) \leqq \frac{1}{2}$ for every S in the set A.

Let $v(P)$ be a finite linear combination, with positive coefficients, of $K(S, P)$'s with S in A:

$$v(P) = \sum_1^m c_\nu \cdot K(S_\nu, P) \qquad (c_\nu > 0; S_\nu \text{ in } A). \tag{4.6}$$

We then have

$$\begin{aligned} v^*_{[G]}(P_0) &= \sum_1^m c_\nu \cdot K^*_{[G]}(S_\nu, P_0) \leqq \frac{1}{2} \sum_1^m c_\nu \\ &= \frac{1}{2} \sum_1^m c_\nu \cdot K(S_\nu, P_0) = \frac{1}{2} v(P_0), \end{aligned} \tag{4.7}$$

for any $v(P)$ of the form (4.6).

More generally, suppose that the function $v(P)$ is expressible in the form of an integral:

$$v(P) = \int_A K(M, P) d\mu(e_M). \tag{4.8}$$

Approximation to this integral by means of Riemann sums yields an approximation to the $v(P)$ in (4.8) by a sequence of functions $v_n(P)$ of the form (4.6). Using the last statement in part (e) of Theorem II, §1, and the result of (4.7) for the functions $v_n(P)$, we obtain

$$
(4.9)\qquad \begin{aligned} v_A(P_0) &\leqq \overset{*}{v}_{[G]}(P_0) \leqq \liminf_{n\to\infty} (v_n)^*_{[G]}(P_0) \\ &\leqq \tfrac{1}{2} \lim_{n\to\infty} v_n(P_0) = \tfrac{1}{2} v(P_0), \end{aligned}
$$

for any $v(P)$ of the form (4.8).

In particular, by Theorem II, §3, $u_A(P)$ is of the form (4.8). Hence, we have

$$u_A(P_0) = (u_A)_A(P_0) \leqq \tfrac{1}{2} u_A(P_0);$$

from which it follows that $u_A(P_0) = 0$, thus that $u_A(P) \equiv 0$.

Lemma 3. *Let $u(P)$ be positive harmonic in D, and let ϵ be an arbitrary positive number. Then there exists a closed subset A of Δ_1 such that $u(P_0) \leqq u_A(P_0) + \epsilon$. The set A depends, of course, upon $u(P)$ and ϵ.*

Denote by $\Gamma_{m,n}$ (m, $n = 1, 2, \cdots$) the set of points of Δ whose distance from Γ_n does not exceed $1/m$. The sets $\Gamma_{m,n}$ are closed, and, for n fixed, $m \to \infty$, we have $\Gamma_{m,n} \downarrow \Gamma_n$. As a consequence of Theorem I, (g), §3, and the result of the preceding lemma, we may, for each n, choose $m = m(n)$ so great that if $B_n = \Gamma_{m(n),n}$, then $u_{B_n}(P_0) < 2^{-n} \cdot \epsilon$. Having for each n selected B_n in this fashion, define C_n as $B_1 + B_2 + \cdots + B_n$. The sets C_n are closed and form with increasing n an increasing sequence. Denote by A_n the closure of the complement on C_n in Δ. The distance of the sets A_n and Γ_n is at least $1/m(n)$; thus the A_n, which form a descending sequence, have an intersection A which is closed and, having no point in common with any Γ_n, is a subset of Δ_1.

We show that this A satisfies the requirements of the lemma. By the construction above,

$$u_{C_n}(P_0) \leqq \sum_{1}^{n} u_{B_\nu}(P_0) < \sum_{1}^{n} 2^{-\nu} \cdot \epsilon < \epsilon.$$

Observing that $A_n + C_n = \Delta$ and using (e) and (h) from Theorem I, §3, we obtain from this

$$u(P_0) = u_\Delta(P_0) = u_{(A_n + C_n)}(P_0) \leqq u_{A_n}(P_0) + u_{C_n}(P_0) \leqq u_{A_n}(P_0) + \epsilon.$$

The limiting form of this inequality as n becomes infinite, calculated with the aid of (g) from the theorem just cited, is the inequality of the lemma.

Lemma 4. *Let A and B be closed subsets of Δ having no common point. Assume that B is a subset of Δ_1 and that ϵ is an arbitrary positive number. Then there exists an open set G containing A such that for every S of B, $K^*_{[G]}(S, P_0) < \epsilon$.*

Let $G_1, G_2, \cdots$ be a descending sequence of open sets which have P_0 as an exterior point, which contain A, and whose closures have A as their intersection. If the present lemma were false, we could find for each n a point S_n

of B and a number δ, positive and independent of n, such that $K^*_{[G_n]}(S_n, P_0) \geqq \delta$. We show that this leads to a contradiction.

Using equation (3.4) from the proof of Theorem II, §3, with $u(P)$ and G there as the present $K(S_n, P)$ and G_n, we have

$$K^*_{[G_n]}(S_n, P) = \int_{\bar{G}_n} K(M, P) d\mu_n(e_M),$$

where $\mu_n(e)$ is a distribution over $\bar{G}_n$ whose total mass, calculated by writing $P=P_0$ in the equation, must be between the numbers δ and 1 inclusive. It is now possible to extract a subsequence of the natural numbers such that the corresponding subsequence of these distributions converges to a weak limiting distribution $\mu_0(e)$ over A having a total mass of at least δ. Since B is closed and compact, it is then possible to extract from this subsequence a second, such that the corresponding subsequence of $\{S_n\}$ converges to a point S_0 in B. We may assume that both these extractions have already been performed, so that

$$\begin{aligned} K(S_0, P) &= \lim_{n\to\infty} K(S_n, P) \leqq \limsup_{n\to\infty} K^*_{[G_n]}(S_n, P) \\ &= \lim_{n\to\infty} \int_{\bar{G}_n} K(M, P) d\mu_n(e_M) = \int_A K(M, P) d\mu_0(e_M). \end{aligned}$$

Since the last integral is positive and since $K(S_0, P)$ is minimal, it follows from Corollary 2 that S_0 is in A. This is the desired contradiction.

LEMMA 5. *Let A be a closed subset of Δ, and E a Borel subset of Δ_1 having no point in common with A. Let $u(P)$ be a harmonic function of the form*

$$u(P) = \int_E K(M, P) d\mu(e_M). \tag{4.10}$$

We then have $u_A(P)=0$.

Assume first that $E=B$, where B is as in the statement of Lemma 4. Let $v(P)$ be a finite linear combination with positive coefficients of $K(S, P)$'s with S in B:

$$v(P) = \sum_1^m c_\nu \cdot K(S_\nu, P) \qquad (c_\nu > 0; S_\nu \text{ in } B). \tag{4.11}$$

Now, ϵ being an arbitrary positive number, let G be the G of Lemma 4 corresponding to it for the present A and B. We then have (cf. (4.7) in the proof of Lemma 2)

$$v^*_{[G]}(P_0) = \sum_1^m c_\nu \cdot K^*_{[G]}(S_\nu, P_0) \leqq \epsilon \cdot \sum_1^m c_\nu = \epsilon \cdot \sum_1^m c_\nu \cdot K(S_\nu, P_0) = \epsilon \cdot v(P_0).$$

By the use of Riemann sums, any $u(P)$ of the form (4.10) (with $E=B$) can be approximated by a sequence of functions $v_n(P)$ of the form (4.11). There follows (cf. (4.9) above)

$$u_A(P_0) \leqq \overset{*}{u}_{[G]}(P_0) \leqq \liminf_{n\to\infty} (v_n)^{*}_{[G]}(P_0) \leqq \lim_{n\to\infty} \epsilon\cdot v_n(P_0) = \epsilon\cdot u(P_0).$$

Since ϵ is arbitrary, $u_A(P_0)=0$; hence, $u_A(P)\equiv 0$.

When E is of a more general form, we may write $E=B+C$, where B and C are without common points, B is as above, and C has a μ-mass smaller than a preassigned ϵ. If now $u(P)$ is decomposed into two parts represented by the integral of (4.10) extended over the sets B and C respectively, then, by what has just been proved, the first of these contributes nothing to the value of $u_A(P_0)$, while the second contributes an amount less than ϵ. It then follows as above that $u_A(P)=0$.

THEOREM III. *Every non-negative harmonic function $u(P)$ in D admits of exactly one canonical representation. The canonical distribution $\mu(e)$ representing $u(P)$ is characterized by the relation*

$$u_A(P) = \int_A K(M, P)d\mu(e_M), \tag{4.12}$$

which holds for every closed subset A of Δ.

We prove first the existence of a canonical representation. Let $u(P)$ be non-negative harmonic, ϵ positive, and A the A of Lemma 3 for this $u(P)$ and this ϵ. Consider the decomposition

$$u(P) = u_A(P) + [u(P) - u_A(P)].$$

By the result of Theorem II, §3, the function $u_A(P)$ admits a representation in terms of a distribution whose total mass is in A, thus a canonical representation. The function $u(P)-u_A(P)$ is non-negative harmonic, and its value at P_0, by the inequality of Lemma 3, cannot exceed ϵ. This means: Any non-negative harmonic function in D can be expressed as the sum of two, one of which admits a canonical representation, and the other of which has a value at P_0 smaller than a preassigned positive number.

Let $\epsilon_1, \epsilon_2, \cdots$ be a decreasing sequence of positive numbers having zero as limit. Starting with the given $u(P)$ effect the decomposition $u(P)=u_1(P)+u_1'(P)$, where $u_1(P)$ admits a canonical representation, and where $u_1'(P_0)<\epsilon_1$. Repeat the process for $u_1'(P)$, writing $u_1'(P)=u_2(P)+u_2'(P)$, where $u_2(P)$ has a canonical representation, and where $u_2'(P_0)<\epsilon_2$. Proceeding in this way we obtain recursively a sequence of decompositions:

$$u(P) = u_1(P) + u_1'(P),\quad u_{n-1}'(P) = u_n(P) + u_n'(P) \qquad (n = 2, 3, \cdots),$$

where $u_n(P)$ has a canonical representation, and where $u_n'(P_0) < \epsilon_n$. Combining the first m of these relations, we get

$$u(P) = \sum_1^m u_n(P) + u_m'(P).$$

Since the $u_m'(P)$ form a decreasing sequence of non-negative harmonic functions vanishing in their limit at P_0, the limit vanishes identically and we have

$$u(P) = \sum_1^\infty u_n(P).$$

Now, for each n, let $\mu_n(e)$ be a canonical distribution representing $u_n(P)$. Write

$$\mu(e) = \sum_1^\infty \mu_n(e).$$

Any distribution represented by a partial sum of this series, since it represents a harmonic function which does not exceed $u(P)$, has a total mass not exceeding $u(P_0)$. Hence, the series defines a finite distribution $\mu(e)$, which is obviously canonical. We now have

$$\begin{aligned} u(P) = \sum_1^\infty u_n(P) &= \sum_1^\infty \int_\Delta K(M, P)d\mu_n(e_M) \\ &= \int_\Delta K(M, P)d\mu(e_M), \end{aligned}$$

which completes the existence proof.

We prove the uniqueness by showing that the relation (4.12) holds for *any* canonical distribution $\mu(e)$ representing $u(P)$. This is sufficient, since (4.12), with $P = P_0$, yields $u_A(P_0) = \mu(A)$, and thus shows that $\mu(e)$ is determined for all closed, hence for all Borel, sets in Δ.

Assume thus that $\mu(e)$ is a canonical distribution representing $u(P)$. For brevity we shall write

$$u(E; P) = \int_E K(M, P)d\mu(e_M),$$

when E is any Borel subset of Δ. Since Δ and Δ_1 differ by the μ-null set Δ_0, we have for such E

$$(4.13) \qquad u(E; P) = u(\Delta_1 E; P).$$

Now let A be a closed subset of Δ. From the additivity of the integral as a function of sets there follows

$$u(P) = u(\Delta; P) = u(\Delta_1; P) = u(\Delta_1 A; P) + u(\Delta_1 - A; P).$$

Applying to this result that of Lemma 5 with E there taken as $\Delta_1 - A$ and A taken as the present A, we obtain

$$(4.14)\qquad u_A(P) = u_A(\Delta_1 A; P) + u_A(\Delta_1 - A; P) = u_A(\Delta_1 A; P).$$

Denote by A_n ($n = 1, 2, \cdots$) the set of points of Δ whose distance from A does not exceed $1/n$, and by B_n the closure of the complement of A_n in Δ. A_n and B_n are closed, $A_n + B_n = \Delta$, and B_n has no point in common with A. With the aid of Lemma 5 with E taken as $\Delta_1 A$ and A there taken as the present B_n, we have

$$\begin{aligned} u(\Delta_1 A; P) &= u_\Delta(\Delta_1 A; P) = u_{(A_n + B_n)}(\Delta_1 A; P) \\ &= u_{A_n}(\Delta_1 A; P) + u_{B_n}(\Delta_1 A; P) = u_{A_n}(\Delta_1 A; P) \leqq u(\Delta_1 A; P). \end{aligned}$$

Since $A_n \downarrow A$, the limiting form of this is

$$(4.15)\qquad u_A(\Delta_1 A; P) = u(\Delta_1 A; P).$$

From (4.14), (4.15), and (4.13) follows

$$u_A(P) = u_A(\Delta_1 A; P) = u(\Delta_1 A; P) = u(A; P),$$

which completes the proof.

COROLLARY 3. *The function $u_A(P)$, defined originally for closed subsets of Δ, admits of extension to a completely additive function of Borel sets in Δ.*

COROLLARY 4. *The condition that Δ_0 be void is both necessary and sufficient for the uniqueness in general*([25]) *of the representation of Theorem* III *of the preceding section.*

If Δ_0 is void, all representations are canonical, and thus unique. On the other hand, if S is a point of Δ_0, $K(S, P)$ has at least two representations, viz., its canonical representation and its representation by a unit point-mass at S.

5. **Examples and applications.** We shall first clear up by an example the question of the existence of a domain for which the set Δ_0 is non-void. Since a point of divergence of the representation in §3 from the Poisson-Stieltjes integral formula lies in the (presumable) failure of that representation to be unique, and since much of the complication of §4 occurs on this account, it is desirable to have this example to show that the difficulty is genuine, and not simply the result of an ineffectiveness of the particular analytical devices employed. To show further that the presence of the set Δ_0 is not connected with any necessary complication in the topological structure of the domain, this

([25]) It is a question of interest whether *bounded* positive harmonic functions can have any but canonical representations. This reduces to an investigation of the representations of $l(P) \equiv 1$. It may be remarked that the function $l_A(P)$ plays a role analogous to that of the swept-out mass function $m(e, P)$, and serves as a natural starting point for an investigation of the "Dirichlet problem" associated with the present notion of ideal boundary.

example has been chosen so that the domain together with its (ordinary) boundary is the topological image of a closed sphere.

We require certain properties of the swept-out mass for a domain limited by a simple closed *surface of bounded curvature*[26]. Such a domain may for our present purposes be characterized by the existence of a positive number r (called here an *admissible radius* for the domain) such that each boundary point of the domain lies at the point of tangency of two spheres of radius r having their interiors respectively interior and exterior to the domain. The boundary surface of such a domain necessarily has in each point a well defined normal varying continuously in direction from point to point, and has an area given by the elementary formula. The distribution $m(e, P)$ is absolutely continuous with respect to area, and its superficial density $\Omega(S, P)$ at a boundary point S is positive, varies continuously with S, and is given by the Poisson formula

$$\Omega(S, P) = \frac{1}{4\pi} \cdot \frac{\partial}{\partial n_S} G(S, P),$$

in which n_S denotes the inward directed normal at the point S, and $G(M, P)$ is the Green's function. $\Omega(S, P)$ is positive harmonic in P, approaches zero as P approaches a boundary point distinct from S, and admits the estimate

$$\Omega(S, P) \leqq \frac{1}{2\pi} \frac{\cos \phi}{\overline{SP}^2} + \frac{1}{4\pi r \cdot \overline{SP}}, \tag{5.1}$$

where ϕ is the angle between the inward normal at S and the directed segment SP, and where r is an admissible radius for the domain.

LEMMA 1. *Let D be a (bounded) domain whose boundary d is of bounded curvature, and let δ be a positive number. There exist positive constants k_1 and k_2 such that*

$$k_1 \cdot \text{dist } (P, d) \leqq \Omega(S, P) \leqq k_2 \cdot \text{dist } (P, d),$$

whenever $\overline{SP} \geqq \delta$. k_1 depends only upon the domain; k_2 depends only upon the domain and δ.

Let r be an admissible radius for D. Let D_1 be the set of points of D whose distance from d is at least r. $\Omega(S, P)$ has a positive lower bound m taken over all P in D_1, S in d. Obviously, if P is in D_1, $\Omega(S, P) \geqq (m/\text{diam } (D)) \cdot \text{dist } (P, d)$. Suppose that $P = P_1$ is a point of D not in D_1. Let S_1 be the closest point of d to P_1. Let Σ_1 be the sphere of radius r internally tangent to d at S_1, and let Q_1

(26) See de la Vallée Poussin, loc. cit.(1). The definition used here is equivalent to that of the cited reference for domains with simple boundaries. It may be remarked, however, that the largest admissible radius in the sense used here is not necessarily identical with the minimum radius of curvature of the boundary surface.

be its center. Q_1 is in D_1, and Q_1, P_1, S_1 lie in that order along the inward normal at S_1. Using the inequality $\Omega(S, Q_1) \geqq m$, and the Harnack inequality for the sphere Σ_1, we obtain $\Omega(S, P_1) \geqq (m/4r)\cdot\overline{P_1S_1} = (m/4r)\cdot\text{dist }(P_1, d)$. Thus we may choose k_1 as the smaller of the two numbers $m/4r$ and $m/\text{diam }(D)$.

For the existence of k_2, let M be an upper bound, estimated by (5.1), of $\Omega(S, P)$ for $\overline{SP} \geqq \delta/3$. If dist $(P, d) \geqq \delta/3$, then the inequality of the lemma is satisfied with $k_2 = 3M/\delta$. Suppose that $P = P_1$ is a point of D such that dist $(P_1, d) < \delta/3$. Let S_1 be a closest point of d to P_1. If now $\overline{S_1P} \leqq \delta/3$ and $\overline{SP_1} \geqq \delta$, then we have $\overline{SP} \geqq \delta/3$, and hence $\Omega(S, P) \leqq M$. Let Σ_1' be the sphere of radius r and center Q_1' externally tangent to d at S_1; let Q_1'' be the midpoint of the segment S_1Q_1'. The function $\phi(S_1; P) = 1 - (\overline{S_1Q_1''}/\overline{PQ_1''})$ is a barrier[27] for D at S_1. The greatest lower bound m' of $\phi(S_1; P)$ for $\overline{PQ_1'} \geqq r$, $\overline{S_1P} = \delta/3$, is positive and depends only upon r and δ. From the properties of a barrier and the relations $\overline{P_1S_1} < \delta/3$ and $\overline{P_1Q_1''} = \overline{P_1S_1} + \overline{S_1Q_1''}$ there follows for $\overline{SP_1} \geqq \delta$, $\Omega(S, P_1) \leqq (2M/m'r)\cdot\overline{S_1P_1}$. We now take k_2 as the larger of the two numbers $2M/m'r$ and $3M/\delta$.

Assume now that a domain D is the sum of two, D_1 and D_2, of which the latter is limited by a surface of bounded curvature. Consider the set σ of boundary points of D_2 at a positive distance from D_1 and at a distance exceeding the positive number δ from that part τ of the boundary of D_1 lying in D_2. We assume that σ is not void; thus, since it is open in the boundary of D_2, it has positive superficial measure. If now a positive mass lying in D_1 is swept out of D, the result is equivalent to that of commencing with D_1 and sweeping out D_1 and D_2 alternately in infinite succession. We seek a bound for the superficial density of the final distribution in points of σ.

LEMMA 2. *Under the circumstances just mentioned, the distribution resulting from sweeping out D has in each point of σ a superficial density which varies continuously with the point and which nowhere in σ exceeds $k\mu/A$, where μ is the total mass lying on τ after the first sweeping-out of D_1, where A is the superficial measure of σ, and where k is a constant which depends only upon the domain D_2 and the number δ.*

Let $f(S)$ be a function of the form

$$f(S) = \int_\tau \Omega_2(S, P)d\mu(e_P), \tag{5.2}$$

where $\mu(e)$ is a distribution of positive mass over τ and $\Omega_2(S, P)$ is $\Omega(S, P)$ for the domain D_2. $f(S)$ is a continuous function of S in σ, and represents the density in the point S of the distribution which results if we sweep the distribution $\mu(e)$ out of D_2. Now let $k = k_2/k_1$, where k_1 and k_2 are from Lemma 1 for D_2 and the present δ. It follows readily from Lemma 1 and (5.2) that

(27) Kellogg, loc. cit., p. 329.

$f(S) \leqq k \cdot f(S')$ for any two points S and S' of σ. Thus, if we denote by $\bar{f}$ and $\underline{f}$ respectively the least upper and greatest lower bounds of $f(S)$ in σ, we have $\bar{f} \leqq k \cdot \underline{f}$.

Now let $f_n(S)$ be the density of the contribution to the mass on σ made by the nth sweeping-out of D_2, and let $g_n(S)$ be the density of the total distribution on σ at this stage. Since the total mass on σ at no stage exceeds μ, we have

$$\sum_1^n f_\nu(S) \leqq \sum_1^n \bar{f}_\nu \leqq \sum_1^n k \cdot \underline{f}_\nu \leqq k \cdot \underline{g}_n = \frac{k}{A} \cdot \int_\sigma \underline{g}_n dS \leqq \frac{k}{A} \cdot \int_\sigma g_n(S) \cdot dS \leqq \frac{k\mu}{A} \cdot$$

This relation not only shows that the infinite sum of the $f_n(S)$ representing the density of the final distribution is convergent and admits the desired upper bound, but also, since the sum of the $\bar{f}_n$ admits the same bound, that the convergence is uniform in σ.

Example 1. *Domain with singular edge.* Let x, y, z be rectangular cartesian coordinates in space. We denote by $C(r)$, where $r > 0$, the capsule-formed domain described by the conditions

$$x^2 + y^2 < r^2; \qquad |z| < 1 + (r^2 - x^2 - y^2)^{1/2}.$$

The surface bounding $C(r)$ evidently has bounded curvature. The segment of the z axis for which $|z| \leqq 1$ we shall refer to as the *core* of $C(r)$. For $\alpha > 0$, we shall denote by $C(r, \alpha)$ the configuration obtained from $C(r)$ by a translation α units in the direction of the positive x axis.

Consider for fixed r, r' the intersection of the closures of the domains $C(r)$ and $C(r', \alpha)$. If $\alpha < r + r'$, this is non-void, and, when $\alpha \uparrow r + r'$, it closes down upon a line segment. Thus it is possible to choose $\alpha < r + r'$ so that the capacity of the intersection is arbitrarily small.

Now let $r_0, r_1, r_2, \cdots$ and $h_1, h_2, \cdots$ be respectively decreasing and increasing sequences of positive numbers of fixed selection subject only to the requirement that $r_0 = 1$, that the sum of the r_n be convergent, and that $h_n \to \infty$. In terms of these sequences we define ($n = 1, 2, \cdots$):

m_n as the minimum value of the function $\Omega(S, P)$ for the domain $C(r_n)$ when S ranges over its boundary and P over its core,

A_n as the area of the surface bounding $C(r_n)$,

k_n as the k of Lemma 2 with $C(r_n)$ for D_2 and $\frac{1}{2}r_n$ for δ,

α_n as a positive number chosen so that $r_{n-1} + \frac{1}{2}r_n < \alpha_n < r_{n-1} + r_n$ and so that the capacity of the intersection of the closures of $C(r_{n-1})$ and $C(r_n, \alpha_n)$ is less than $m_n A_n / 2k_n h_n$,

β_n as $\alpha_1 + \alpha_2 + \cdots + \alpha_n$,

C_0 as $C(r_0) = C(1)$, and C_n as $C(r_n, \beta_n)$,

D_0 as C_0, and D_n as $C_0 + C_1 + \cdots + C_n$.

The domains $C_0, C_1, C_2, \cdots$ are decreasing in size and lie parallel to each other with their centers in order along the positive x axis. Two successive do-

mains intersect each other, but neither contains the center of the other. The centers have as limit point the point $(\beta, 0, 0)$, where $\beta = \lim \beta_n < 2\sum r_n$.

We now define the domain D as $\lim D_n$. D is bounded, and together with its boundary it is the topological image of a closed sphere. Every boundary point is regular with respect to the Dirichlet problem; in fact, Poincaré's "cone condition" applies in every boundary point. The boundary consists of parts of the boundaries of the C_n and also the limiting line segment $x = \beta$, $y = 0$, $|z| \leqq 1$. This segment we shall refer to as the *singular edge* of D. It is convenient to take for P_0 the center of C_0, that is, the origin of coordinates.

THEOREM I. *Any function $u(P)$ positive harmonic in D_n taking on continuous boundary values which are zero for those boundary points not in C_{n+1} satisfies*

$$u(P) \geqq h_n \cdot u(P_0) \tag{5.3}$$

for all P on the core of C_n. Thus, if $u(P)$ is positive harmonic in D and approaches zero at every boundary point not on the singular edge, it is unbounded in the neighborhood of every point of the singular edge([28]).

Consider D_n as $D_{n-1} + C_n$ $(n \geqq 1)$. Consider the set σ_n of boundary points of C_n whose distance from that part τ_n of the boundary of D'_{n-1} in C_n exceeds $\frac{1}{2}r_n$. Since $\alpha_n > r_{n-1} + \frac{1}{2}r_n$, the set σ_n includes all boundary points of C_n whose x-coordinate exceeds β_n; thus, it has a superficial measure of at least $\frac{1}{2}A_n$. Since the distance of τ_n from P_0 is at least unity, the total mass received by τ_n when a unit mass at P_0 is swept out of D_{n-1} cannot exceed the capacity of τ_n, thus cannot exceed $m_n A_n / 2h_n k_n$. It follows from Lemma 2 that the density in points of σ_n of the distribution resulting when a unit mass at P_0 is swept out of D_n cannot exceed m_n/h_n. On the other hand, when a unit mass at P on the core of C_n is swept out of C_n, the resulting density at points of σ_n is at least m_n; this remains true *a fortiori* if the sweeping-out is continued into D_n. Since the boundary points of D_n in C_{n+1} are in σ_n, the relation (5.3) is immediate from the estimates just made if $u(P)$ is represented in D_n in terms of its boundary values by means of (1.1). The last statement of the theorem is an obvious corollary.

([28]) This example, here introduced as auxiliary to the construction of Example 2, is not without intrinsic interest. First, it exhibits another way, besides that already pointed out by Bouligand, in which the principle of Picard may be in default. For, consider a point Q of the singular edge. Instead of having at least two linearly independent positive harmonic functions approaching zero at all boundary points except Q, we have none at all. Second, with a suitable choice of the constants, the example serves to answer in the negative a conjecture of N. Wiener, *Discontinuous boundary conditions and the Dirichlet problem*, these Transactions, vol. 25 (1923), p. 313. For, if the numbers h_n are so chosen that, say, $h_n/2^n \to \infty$, then it is possible to define a non-negative, summable boundary function $\phi(Q)$ which vanishes at all boundary points Q whose z-coordinate is less than unity, and which determines a harmonic function unbounded near all points of the singular edge. Thus, in particular, we have an example of a simple domain for which the condition (γ), Maria and Martin, loc. cit., p. 519, is not fulfilled.

Example 2. *Domain for which Δ_0 is non-void.* We write D_2 for the D of Example 1. Retaining the notations introduced in connection with that example, let A and A' be the end-points of the singular edge, and let Σ and Σ' be open spheres of radius $\frac{1}{2}$ (<1) having these points respectively for centers([29]). We define the domain D as $D_2+\Sigma+\Sigma'$. This domain is symmetric about the plane $z=0$, it has the same simple connectivity as does D_2, and it is regular with respect to the Dirichlet problem. It is convenient also to define the auxiliary domain $D_1=D_2-\overline{\Sigma}-\overline{\Sigma}'$. Since $D_1\subset D_2\subset D$, we have

$$G_1(M, P) < G_2(M, P) < G(M, P)$$

for the corresponding Green's functions.

Form the functions $K_2(M, P)$ and $K(M, P)$ corresponding respectively to D_2 and D as in §2, taking in both cases P_0 as the origin of coordinates. Consider a sequence $\{M_n\}$ of points which lie on the x axis and whose x coordinates form an increasing sequence with limit β. Without loss of generality it may be assumed that $\{M_n\}$ is fundamental for both D_2 and D, thus determining for them ideal boundary elements S_2 and S together with the corresponding harmonic functions $K_2(S_2, P)$ and $K(S, P)$. We shall prove that $K(S, P)$ is *not* minimal for D.

In D_1 we have

$$G(M, P) = G_1(M, P) + G^*_{[\Sigma+\Sigma']}(M, P).$$

From this, on writing $v_n(P)=K^*_{[\Sigma+\Sigma']}(M_n, P)$ and $w_n(P)=G_1(M_n,P)/G(M_n, P_0)$,

$$K(M_n, P) = w_n(P) + v_n(P).$$

Now $w_n(P)\to 0$ in D_1. If not, a subsequence would have a positive harmonic limit $w(P)$ in D_1. If k were any positive number less than $w(P_0)$, we should have infinitely many n such that $w_n(P_0)=G_1(M_n, P_0)/G(M_n, P_0)>k$. This would imply that

$$\begin{aligned} K_2(S_2, P) &= \lim_{n\to\infty} \frac{G_2(M_n, P)}{G_2(M_n, P_0)} \leqq \liminf_{n\to\infty} \frac{G(M_n, P)}{G_1(M_n, P_0)} \\ &\leqq \limsup_{n\to\infty} \frac{1}{k}\cdot\frac{G(M_n, P)}{G(M_n, P_0)} = \frac{1}{k}\cdot K(S, P). \end{aligned}$$

But this is impossible since, in particular, $K(S, P)$ is harmonic at $P=A$, while $K_2(S_2, P)$, as follows readily from Theorem I, is unbounded near A. Thus we have proved that $v_n(P)\to K(S, P)$ for P in D_1.

The functions $v_n(P)$ approach zero in every boundary point of D; further-

---

([29]) The device of forming two modified domains by means of a sphere (here two spheres) is due to A. J. Maria. The argument below which enables us to prove the boundedness of $K(S, P)$ at points distant from $\Sigma+\Sigma'$ is a paraphrase of that shown the author by Maria some time ago.

more, they are uniformly bounded over any set at a positive distance from $\Sigma+\Sigma'$. In fact,

$$v_n(P) \leqq K^*_{[\Sigma]}(M_n, P) + K^*_{[\Sigma']}(M_n, P),$$

and, for $\overline{AP} > \frac{1}{2}$, we have[30]

$$K^*_{[\Sigma]}(M_n, P) \leqq K(M_n, A) \cdot \frac{2 \cdot \overline{AP} + 1}{(2 \cdot \overline{AP} - 1)^2},$$

together with an analogous bound for the other function when $\overline{A'P} > \frac{1}{2}$. This result, coupled with the result of the preceding paragraph, implies that $K(S, P)$ approaches zero in every boundary point of D which is at a positive distance from the spheres Σ and Σ'.

Now let σ be the closure of the set of those boundary points of D which are limit points of Σ and for which $K(S, P)$ has a positive superior limit[31]. Similarly define σ' in terms of Σ'. Since $K(S, P)$ is symmetrical about the xy plane, σ and σ' are reflections of each other in it. Let σ_n and σ_n' be the sets of points of D whose distance from σ and σ' respectively does not exceed $1/n$. In $D-\sigma_n-\sigma_n'$, $K(S, P)$ is bounded and takes on continuous boundary values. These are zero at those boundary points which are also boundary points of D. Thus,

$$K(S, P) = K^*_{(\sigma_n+\sigma_n')}(S, P). \tag{5.4}$$

Write $u_n(P) = K^*_{\sigma_n}(S, P)$. If $n > 1$, the boundary values determining this function in $D-\sigma_n$ are continuous; hence the function itself must approach zero in every boundary point of D not in σ. Since as $n \to \infty$ the $u_n(P)$ form a descending sequence, they have a limit $u(P)$ which is non-negative harmonic in D and which approaches zero in every boundary point not in σ. An analogous statement holds for the functions $u_n'(P)$ similarly defined in terms of σ_n' and for their limit $u'(P)$. By the symmetry of the construction, the functions $u(P)$ and $u'(P)$ are images of each other in the xy plane. Neither function can be zero, since (5.4) in conjunction with (h) of Theorem II, §1, shows that $K(S, P)$ is dominated by $u_n(P)+u_n'(P)$ and hence by $u(P)+u'(P)$. Now $K(S, P)$ dominates both $u(P)$ and $u'(P)$. Thus, if $K(S, P)$ were minimal, these functions would be multiples of each other. This would give a contradiction, since, by what was shown above, it would imply that both functions approach zero in every boundary point of D. This proves

THEOREM II. *There exist structurally simple domains for which the set* Δ_0 *is not vacuous.*

---

(30) This is readily shown by considering a smaller sphere Σ_1 concentric with Σ, solving the exterior Dirichlet problem for Σ_1 with boundary values $\phi(Q) = K(M_n, Q)$, and making $\Sigma_1 \uparrow \Sigma$.

(31) The set σ actually consists of the single point in which the surface of Σ intersects the singular edge of D_2.

Let us turn now to another point. The definition of ideal boundary point in §2 is purely potential-theoretic in the sense that it involves the actual structure of the domain only indirectly as the structure influences the behavior of Green's functions. It is natural to ask whether these boundary elements are identifiable with some suitable topologically defined boundary elements analogous, perhaps, to prime-ends([32]). We show by Example 3 that if by this there is meant a *purely* topological definition, that is, naturally, one which is invariant under topological mappings of two domains together with their boundaries upon each other, then the answer is in the negative. This, of course, does not preclude the possibility of obtaining a more geometric definition equivalent to that of §2, but indicates, rather, that such a definition must take into account the metric structure of the domain as well. The domain is of sufficient simplicity to permit the carrying out explicitly of the representation of the positive harmonic functions in the terms of the present analysis, and we do so.

Example 3. *A simple domain in which a certain boundary point corresponds to a continuum of ideal boundary elements.* The domain we consider is one of the two bounded by the surface of two spheres, one internally tangent to the other, and by a plane containing the common diameter. The boundary point of interest is naturally the point of tangency of the bounding spheres. By a succession of inversions commencing with an inversion about this point the domain can be thrown into([33])

$$D:\quad 0 < x < \infty, \qquad -\infty < y < \infty, \qquad -\tfrac{1}{2}\pi < z < \tfrac{1}{2}\pi.$$

The corresponding Kelvin transformation reduces the study of the positive harmonic functions of the original domain to the study of those of D. Evidently the minimal property is preserved under the transformation.

For this discussion we shall designate (ξ, η, ζ) as the coordinates of a point M of D and by (x, y, z) those of P. (ρ, ϕ, ζ) and (r, θ, z), where $\xi = \rho\cos\phi$, etc., will be the corresponding cylindrical coordinates. For P_0 we choose $(1, 0, 0)$.

By the methods of images, the Green's function $G(M, P)$ for D may be determined as a conditionally convergent lattice potential:

$$G(M, P) = \sum_{\mu=-\infty}^{\infty} \sum_{\nu=0}^{1} (-1)^{\mu+\nu}\,\overline{(M)_{\mu,\nu}P}^{-1}, \tag{5.5}$$

in which $(M)_{\mu,\nu}$ denotes the point $[(-1)^\nu\xi,\ \eta,\ \mu\pi+(-1)^\mu\zeta]$.

([32]) In particular, we may cite the definition of boundary element given by F. W. Perkins, *The Dirichlet problem for domains with multiple boundary points*, these Transactions, vol. 38 (1935), pp. 106–144. In so far as the results of Green, loc. cit., are dependent upon this definition of boundary element, there will necessarily be a point of divergence from the present analysis.

([33]) Exactly this domain D has been considered from a closely allied viewpoint of Bouligand, loc. cit. (*Étude des singularités*, $\cdots$), pp. 140–144. Certain of the facts here derived are contained in his work.

We form the function $K(M, P)$ and suppose that M runs through a sequence having no point of accumulation in D. Two cases are to be distinguished: (1) $M \to S$, where S is a boundary point of D; (2) $\rho \to \infty$ and $\phi \to \alpha$ where α is some angle between $-\frac{1}{2}\pi$ and $\frac{1}{2}\pi$ inclusive.

In the first case, $K(M, P)$ approaches a harmonic limit in D which depends upon S but not on the particular manner of approach. The limit function approaches zero as P approaches any boundary point of D distinct from S and also approaches zero at infinity. To see this, we may observe that the series in (5.5) extends the definition of $G(M, P)$ to the whole of space. In particular, when M does not coincide with P or any of its image points, $G(M, P)$ is analytic in ξ, η, and ζ, of period 2π in ζ, odd in $\zeta \pm \frac{1}{2}\pi$, and odd in ξ; similar statements hold in x, y, and z. From this it follows that if we write $h(M) = \xi \cos \zeta$, the quotient $G(M, P)/h(M)$ has a determinate limit $H(S, P)$ as $M \to S$. Writing (ξ_0, η_0, ζ_0) for the coordinates of S, we have

$$H(S, P) = \begin{cases} \left(\dfrac{\partial G}{\partial \xi} \Big/ \dfrac{\partial h}{\partial \xi}\right)_{M=S} & (\xi_0 = 0;\ \zeta_0 \neq \pm \frac{1}{2}\pi), \\[2ex] \left(\dfrac{\partial G}{\partial \zeta} \Big/ \dfrac{\partial h}{\partial \zeta}\right)_{M=S} & (\xi_0 \neq 0;\ \zeta_0 = \pm \frac{1}{2}\pi), \\[2ex] \left(\dfrac{\partial^2 G}{\partial \xi \partial \zeta} \Big/ \dfrac{\partial^2 h}{\partial \xi \partial \zeta}\right)_{M=S} & (\xi_0 = 0;\ \zeta_0 = \pm \frac{1}{2}\pi). \end{cases}$$

A straightforward computation of these derivatives in (5.5) shows that $H(S, P)$, as a function of P, has singularities at S and its image points and only there; furthermore that it approaches zero as P moves to infinity in D. Because of the presence of these singularities, $H(S, P)$ cannot vanish identically. In particular, $H(S, P)$ must be positive harmonic in D. Since the parity relations of $G(M, P)$ in x and z are preserved in $H(S, P)$, the latter function must vanish in every boundary point of D distinct from S. Finally,

$$\frac{H(S, P)}{H(S, P_0)} = \lim_{M \to S} \frac{G(M, P)/h(M)}{G(M, P_0)/h(M)} = \lim_{M \to S} K(M, P).$$

In the second case, $K(M, P)$ approaches a positive harmonic limit which depends upon α but is otherwise independent of the manner of approach of M to infinity. The limit function, which approaches zero in every boundary point of D but is unbounded at infinity, is given explicitly by

$$\text{(5.6)} \qquad \begin{array}{ll} \operatorname{csch}(\cos \alpha) \cdot \cos z \cdot \exp(y \sin \alpha) \cdot \sinh(x \cos \alpha) & (|\alpha| < \frac{1}{2}\pi), \\ x \cdot \cos z \cdot \exp(y \sin \alpha) & (\alpha = \pm \frac{1}{2}\pi). \end{array}$$

This result can be derived from an asymptotic development for $G(M, P)$ valid as M recedes to infinity. To obtain such an expression from (5.5) we apply contour integration to each of the functions

$$\cos \zeta [\sin \zeta - \sin t]^{-1}[(x \pm \xi)^2 + (y - \eta)^2 + (z - t)^2]^{-1/2}$$

made single valued in the complex t plane by cuts parallel to the imaginary axis from the branch points to infinity, the integration in each case being around the boundary of the cut plane. The result is

$$G(M, P) = \int_0^\infty \left\{ \frac{\phi(\zeta, z; \lambda + p)}{(\lambda^2 + 2\lambda p)^{1/2}} - \frac{\phi(\zeta, z; \lambda + q)}{(\lambda^2 + 2\lambda q)^{1/2}} \right\} d\lambda, \tag{5.7}$$

where

$$p = [(x - \xi)^2 + (y - \eta)^2]^{1/2}, \quad q = [(x + \xi)^2 + (y - \eta)^2]^{1/2},$$

$$\phi(\zeta, z; \lambda) = \frac{2}{\pi} \cdot \frac{\cos \zeta \cdot \cos z \cdot \sinh \lambda}{[\cosh \lambda + \cos (\zeta + z)][\cosh \lambda - \cos (\zeta - z)]}.$$

Now evidently as $\lambda \to \infty$ we have, uniformly in ζ and z, $\phi(\zeta, z; \lambda) \sim (4/\pi) \cos \zeta \cdot \cos z \cdot e^{-\lambda}$, where by this is meant that the quotient of the two functions approaches unity. Also, as $\rho \to \infty$, uniformly for r bounded and $\lambda > 0$,

$$p = \rho - r \cos (\theta - \phi) + O(\rho^{-1}); \quad q = \rho + r \cos (\theta + \phi) + O(\rho^{-1});$$
$$(\lambda^2 + 2\lambda p)^{1/2} \sim (\lambda^2 + 2\lambda q)^{1/2} \sim (\lambda^2 + 2\lambda \rho)^{1/2}.$$

From these and (5.7) we have as $\rho \to \infty$, uniformly for r bounded,

$$G(M, P) \sim \cos \zeta \cdot \cos z \cdot \{\exp [r \cos (\theta - \phi)] - \exp [-r \cos (\theta - \phi)]\} \cdot \Phi(\rho),$$

in which

$$\Phi(\rho) = \frac{4}{\pi} \int_0^\infty \frac{e^{-(\lambda + \rho)}}{(\lambda^2 + 2\lambda\rho)^{1/2}} d\lambda.$$

The expressions in (5.6) are now immediate.

Since a sequence of points M having no point of accumulation in D has a subsequence falling under one or the other of the above two cases, it follows from the results of the preceding two paragraphs that these cases subsume all fundamental sequences. Thus the ideal boundary elements of D fall into two corresponding classes. Those of the first class may be identified with boundary points of D, the relative neighborhoods in D being the same as those of the corresponding boundary point. The same symbol will be used to designate a boundary point of D and its corresponding boundary element of this class. The ideal boundary elements of the second class are to be identified with directions α of approach to infinity. We denote these elements by S_α, where $-\frac{1}{2}\pi \leqq \alpha \leqq \frac{1}{2}\pi$. The sets $N(R, \beta, \gamma)$ consisting of all P in D for which $r > R$ and $\beta < \theta < \gamma$, where $R \geqq 0$ and $-\frac{1}{2}\pi \leqq \beta < \gamma \leqq \frac{1}{2}\pi$ form a convenient system of relative neighborhoods in D of the boundary elements S_α.

It is now readily proved that the set Δ_0 for D is void. First, let S be a boundary element of the first class, and let σ be the closure in D of a relative

neighborhood of S. By an argument similar to that used to establish (5.4), we obtain $K_\sigma^*(S, P) = K(S, P)$; from this follows $\psi(S) = 1$. To prove that the functions $K(S_\alpha, P)$ are minimal, assume first that $-\frac{1}{2}\pi < \alpha < \frac{1}{2}\pi$. Let $\sigma(R, \beta, \gamma)$ be the closure in D of the set $N(R, \beta, \gamma)$ above. It is sufficient to show that $F(\alpha, R, \beta, \gamma) \equiv K^*_{\sigma(R,\beta,\gamma)}(S_\alpha, P_0) = 1$ whenever $\beta < \alpha < \gamma$. That the latter is true when $\beta = -\frac{1}{2}\pi$ and $\gamma = \frac{1}{2}\pi$ is easily shown by another argument similar to that cited above from (5.4). Using (h), Theorem II, §1, we then have

$$\begin{aligned} F(\alpha, R, -\tfrac{1}{2}\pi, \beta) + F(\alpha, R, \beta, \gamma) + F(\alpha, R, \gamma, \tfrac{1}{2}\pi) \\ \geqq F(\alpha, R, -\tfrac{1}{2}\pi, \tfrac{1}{2}\pi) = 1. \end{aligned} \tag{5.8}$$

Since F is a positive non-increasing function of R whose value never exceeds unity, it suffices to show that the first and third terms of the first member of (5.8) approach zero as $R \to \infty$. Consider, for example, $F(\alpha, R, -\frac{1}{2}\pi, \beta)$. Let $M(R)$ be the maximum value of the quotient $K(S_\alpha, P)/K(S_\beta, P)$ for P in $\sigma = \sigma(R, -\frac{1}{2}\pi, \beta)$. By a direct estimate from (5.6) we have $M(R) = O(\exp[R \cdot \cos(\alpha - \beta) - R]) \to 0$ as $R \to \infty$. Also $F(\alpha, R, -\frac{1}{2}\pi, \beta) = K_\sigma^*(S_\alpha, P_0) \leqq M(R) \cdot K_\sigma^*(S_\beta, P_0) \leqq M(R)$. This completes the proof for $-\frac{1}{2}\pi < \alpha < \frac{1}{2}\pi$. If α has one of the extreme values, say $\frac{1}{2}\pi$, we write $\gamma = \frac{1}{2}\pi$, and omit the third term of the first member of (5.8). The rest of the proof proceeds as before.

Now return to the original domain bounded by the two spheres and the plane. Denote this domain by D', and by T the point of tangency of the two bounding spheres. If P_0' is the image of P_0 under the conformal transformation which maps D on D', then, as is readily verified, the functions $K'(M, P)$ formed for D' with normalization at P_0' are fixed multiples of the Kelvin transforms of the functions $K(M, P)$ for D. Thus the neighborhoods in D' of its ideal boundary elements are the images under the mapping of the corresponding neighborhoods in D. In particular, neighborhoods in D of boundary elements of the first class map into ordinary neighborhoods in D' of its boundary points distinct from T; the neighborhoods $N(R, \beta, \gamma)$ of the S_α map into fang-shaped sets in D' each having at its apex the point T. We denote by T_α the ideal boundary element in D' corresponding to S_α in D. The set of all T_α $(-\frac{1}{2}\pi \leqq \alpha \leqq \frac{1}{2}\pi)$ is evidently closed in the ideal boundary of D'. The general representation of §3 coupled with the uniqueness theorem of §4 now gives the following representation for the most general positive harmonic in D':

$$u(P) = \int_{d'-\{T\}} K'(Q, P)\,d\mu(e_Q) + \int_{-\pi/2}^{\pi/2} K'(T_\alpha, P)\,d\nu(e_\alpha),$$

where $\mu(e)$ is a non-negative distribution over $d' - \{T\}$ and $\nu(e)$ is a similar distribution over the interval $(-\frac{1}{2}\pi, \frac{1}{2}\pi)$. $\mu(e)$ and $\nu(e)$ are both uniquely determined by $u(P)$.

The University of Illinois,
Urbana, Ill.

In Memoriam

William Caspar Graustein

Born November 15, 1888
Died January 22, 1941

Editor of the Transactions
1936-1941

ON THE FOUNDATIONS OF CALCULUS OF VARIATIONS

BY

HERBERT BUSEMANN AND WALTHER MAYER

The subject of this paper will be variational problems $\int F(x, \dot{x})dt = \min$ in parameter form with fixed endpoints. The existence of rectifiable minimizing arcs has been proved under exceedingly general conditions. However, as soon as one wants to establish differentiability properties of the solutions one uses the Euler equations and must therefore assume the existence of second partial derivatives of $F(x, \dot{x})$.

Hence it is not at all clear *exactly which differentiability properties of the solutions are due to which properties of* $F(x, \dot{x})$. The present paper tries to take a first step towards filling this gap. Simple examples show that without continuity or without the strict convexity of the indicatrix of $F(x, \dot{x})$ no general statements about the differentiability of the solution will be possible. Also, an example was given(1) to show that even if these two conditions are satisfied, the minimizing curves are not necessarily of class D'. In the example the indicatrix is an ellipse everywhere so that the variation of $F(x, \dot{x})$ for fixed x is as smooth as possible. This and the example of the Minkowskian geometry (corresponding to an integrand not depending on the x) suggest investigation of the implications of the *Lipschitz condition*:

$$|F(\bar{x}, \xi) - F(x, \xi)| \leqq C[\sum (\bar{x}_i - x_i)^2]^{1/2}[\sum (\xi^i)^2]^{1/2}.$$

If the underlying space is *two-dimensional* we shall prove that *already under this condition all minimizing arcs are continuously differentiable.* Since we do not even require the existence of first partial derivatives of $F(x, \dot{x})$, our method has to be quite different from the usual ones. Our procedure was suggested by the treatment of geodesics on arbitrary convex surfaces in [4]. So far we have not succeeded in getting a similar result for more dimensional spaces. In the two-dimensional case three questions arise next—namely, whether under the same conditions the minimizing curves have second derivatives, whether every line element is contained in at least one minimizing arc, and whether a given line element occurs in at most one such arc. Simple examples will show that the *answer to all three questions is negative.*

We thought it useful for further investigations of this type to develop the general theory on a broader and more axiomatic base than the results mentioned would have required. Among the results which we get, the following may be of interest. If $F(x, \dot{x}) > 0$ for $\sum(\dot{x}_i)^2 > 0$ and continuous, then

Presented to the Society, September 12, 1940; received by the editors May 22, 1940.

(1) See [3, pp. 281–285]. For the definition of classes C' and D', see §1 below.

$d(P, Q)$ = greatest lower bound of $\int_D F(x, \dot{x})dt$, where D varies over all oriented curves of class D' from P to Q, defines a (not necessarily symmetric) metric space. The question arises how far $F(x, \dot{x})$ is determined by $d(P, Q)$. The answer is: $d(P, Q)$ *does not change if, and only if, the convex closure of the indicatrix of the new function* $F_1(x, \dot{x})$ *at every point coincides with the convex closure of the indicatrix of* $F(x, \dot{x})$. In particular, there is one and only one quasi-regular function $\overline{F}(x, \dot{x})$ to which $d(P, Q)$ belongs.

1. **The space and its vector function.** Our space $\mathfrak{M}$ will be an n-dimensional *coordinate space of class* 1, i.e., it satisfies the following conditions:

(a) It is a Hausdorff space.

(b) Among the neighborhoods of a point P there is at least one coordinate neighborhood (we shall abbreviate this as C.N.) $U(P)$, i.e., there exists a topological mapping $Q \rightarrow x = (x_1, \cdots, x_n)$ of $U(P)$ onto some sphere $\sum_{i=1}^n x_i^2 < \rho$, $\rho > 0$, of the n-dimensional Euclidean space. $U(P)$ together with this mapping forms the C.N. We call $x_1, \cdots, x_n$ the coordinates of Q. This requirement implies: When the intersection $U \cdot \overline{U}$ of two C.N. U and $\overline{U}$ (of the same or different points) is not empty, then each point of $U \cdot \overline{U}$ has two sets of coordinates $x_1, \cdots, x_n$ and $\bar{x}_1, \cdots, \bar{x}_n$. In $U \cdot \overline{U}$ the x_i are single valued and continuous functions of the $\bar{x}_i$ and conversely

$$x_i = f_i(\bar{x}), \qquad \bar{x}_i = \bar{f}_i(x).$$

(c) The functions $f_i(\bar{x})$ and $\bar{f}_i(\bar{x})$ have continuous first partial derivatives.

A *parametrized curve* $\mathfrak{C}$ in $\mathfrak{M}$ is defined in the following way: It is the continuous image of the interval $a \leqq t \leqq b$. It is oriented according to the orientation of (a, b). If a certain subarc $\mathfrak{C}'$: $a' \leqq t \leqq b'$ of $\mathfrak{C}$ is contained in a C.N. the arc $\mathfrak{C}'$ will have a representation $x_i = x_i(t)$, $a' \leqq t \leqq b'$ where the functions $x_i(t)$ are continuous. If the functions $x_i(t)$ have continuous derivatives and if for no value t_0 all $\dot{x}_i(t_0)$ vanish, $\mathfrak{C}'$ will be called of class C', with respect to this C.N. If any subarc $\mathfrak{C}''$ of $\mathfrak{C}'$ is contained in a C.N. U'', the arc $\mathfrak{C}''$ will be of class C' with respect to U''. We can therefore say: We call $\mathfrak{C}$ of *class* C' if every subarc of $\mathfrak{C}$ which is completely in one C.N. is of class C' with respect to this C.N.

Two parametrized curves $\mathfrak{C}$ and $\mathfrak{C}'$, defined for $a \leqq t \leqq b$ and $\alpha \leqq \tau \leqq \beta$ respectively, are said to belong to the *same class* if a topological mapping $t \rightarrow \tau(t)$ of (a, b) onto (α, β) exists which preserves orientation and carries $\mathfrak{C}$ into $\mathfrak{C}'$. A class of parametrized curves is called a *continuous curve.* All curves of a class have the same initial and terminal points, which we therefore call the initial and terminal points of the continuous curve.

If $\mathfrak{C}$ is of class C' and if $\mathfrak{C}$ can be transformed into $\mathfrak{C}'$ by a mapping $t \rightarrow \tau(t)$ where τ is of class C' and $d\tau/dt > 0$ for all t in (a, b), the curve $\mathfrak{C}'$ will be of class C'. The parametrized curves $\mathfrak{C}'$ originating from $\mathfrak{C}$ in this way again form a class. We call this class a *curve of class* C' (or simply a C'-curve). Statements involving the words "continuous curve" or "C'-curve" will always

mean that they hold for all parametrized curves in the corresponding class.

A continuous curve will be called of *class* D' if it is the sum of a finite number of curves of class C'.

Since $\mathfrak{M}$ is of class 1, contravariant vectors are defined. A contravariant vector with origin x and components $\lambda^1, \cdots, \lambda^n$ will be designated by $(x_1, \cdots, x_n; \lambda^1, \cdots, \lambda^n)$ or in short by (x, λ). If $\lambda^j = 0$ for all j we call the vector a null-vector. $\mathfrak{M}$ becomes a *Finsler-manifold* or Finsler-space as soon as a real function $F(x, \lambda)$ is defined for all contravariant vectors and has the following three properties:

(1a) $F(x, \lambda)$ is continuous as a function of the $2n$ variables x_i, λ^i when x_j varies in a C.N.

(1b) $F(x, \lambda)$ is positive definite, i.e., positive if (x, λ) is no null-vector.

(1c) $F(x, \lambda)$ is positively homogeneous of order 1 in λ, i.e., $F(x, c\lambda) = cF(x, \lambda)$ for $c > 0$.

It follows from (a) and (c) that F vanishes for every null-vector. We do not require that $F(x, \lambda) = F(x, -\lambda)$. A neighborhood of P which together with its closure is contained in a C.N. will be called a k-neighborhood (K.N.) of P. From (1a, b, c) we get the following useful

LEMMA 1. *For every K.N. $U(P)$ of P two positive numbers A and B, depending on the coordinate system, can be formed such that*

$$A\,|\lambda| \geqq F(x, \lambda) \geqq B\,|\lambda|, \tag{2}$$

$$A \max_i |\lambda^i| \geqq F(x, \lambda) \geqq B \max_i |\lambda^i|. \tag{2'}$$

$|\lambda^i|$ designates, of course, the absolute value of λ^i; but $|\lambda| = (\sum(\lambda^i)^2)^{1/2}$ and more generally $|a|$ designates $(\sum a_i^2)^{1/2}$ when $a_1, \cdots, a_m$ is a set of numbers with known subscripts.

Let $\overline{C}_{PQ}$ be an oriented parametrized curve of class D' from P to Q, $a \leqq t \leqq b$. Then in each of its points with a finite number of exceptions the contravariant tangential vector $\dot{x}_i(t)$ is defined and $F(x(t), \dot{x}(t))$ will be a piecewise continuous function. Hence

$$I(\overline{C}_{PQ}) = \int_a^b F(x, \dot{x})dt$$

is positive and finite and will have the same value for all parametrized curves which belong to the same class C_{PQ} as $\overline{C}_{PQ}$. We may therefore write $I(C_{PQ})$ instead of $I(\overline{C}_{PQ})$. As F-distance $d(P, Q)$ we define the greatest lower bound of $I(C_{PQ})$ for all oriented curves of class D' from P to Q. We have

(3a) $d(P, Q) > 0$ for $P \neq Q$, $d(P, P) = 0$.

(3b) $d(P, Q) + d(Q, R) \geqq d(P, R)$.

(3c) $d(P, Q)$ is continuous in the product space $\mathfrak{M} \times \mathfrak{M}$.

Finally we require the finite F-compactness of $\mathfrak{M}$:

(1d) If the distances $d(X_\nu, P)$ or $d(P, X_\nu)$ are bounded, then the sequence X_ν, $\nu=1, 2, \cdots$, has an accumulation point.

By $\mathfrak{U}_\rho(P)$ and $\mathfrak{W}_\rho(P)$ we designate the sets of points X with

$$d(P, X) < \rho, \qquad d(X, P) < \rho,$$

respectively. If P and ρ vary, the $\mathfrak{U}_\rho(P)$ will traverse a system of neighborhoods equivalent to the system of all neighborhoods, and so will the $\mathfrak{W}_\rho(P)$.

A curve C_{PQ} from P to Q of class D' will be called minimizing if $I(C_{PQ}) = d(P, Q)$. If R, S, T are any three points which lie in this order on C_{PQ} we have

$$d(R, S) + d(S, T) = d(R, T). \tag{4}$$

An oriented continuous curve from P to Q on which (4) holds for any three points which occur in this order on C_{PQ} will be called a Hilbert curve([2]). A Hilbert curve is necessarily simple (a Jordan arc); therefore we also call it a Hilbert arc. By the well known procedure of successive construction of mid-points one shows that for any pair of points P, Q there exists a Hilbert arc from P to Q.

Let $\overline{C}_{PQ}=P(t)$, $a \leqq t \leqq b$, be any parametrized curve from P to Q, $a=t_0<t_1< \cdots <t_n=b$ a subdivision of (a, b). We form

$$S_m = \sum_{i=0}^{m-1} d(P_i, P_{i+1}) \tag{4'}$$

and call $L(\overline{C}_{PQ})$ the least upper bound of S_m as $(t_0, \cdots, t_m)$ varies over all subdivisions of (a, b). It follows in the usual way from (3a, b, c) that S_m will tend to $L(\overline{C}_{PQ})$ whenever $(t_0, \cdots, t_m)$ traverses a sequence of subdivisions for which $\max_i (t_{i+1}-t_i)$ tends to zero. $L(\overline{C}_{PQ})$ will have the same value for all parametrized curves which are in the same class C_{PQ} as $\overline{C}_{PQ}$. We may therefore write $L(C_{PQ})$ instead of $L(\overline{C}_{PQ})$ which we call the *generalized length* of C_{PQ}. It follows from (4) that for Hilbert arcs H_{PQ}

$$L(H_{PQ}) = d(P, Q). \tag{5}$$

Conversely, if (5) is satisfied, H_{PQ} will be a Hilbert curve.

We conclude this section with a remark which will be useful later. Let the conditions of Lemma 1 be satisfied and let C_{PQ} be a curve of class D' in the K.N. $U(P)$. It follows from (2) that

$$AE(C_{PQ}) \geqq I(C_{PQ}) \geqq BE(C_{PQ}), \tag{6}$$

where $E(C)$ designates generally the Euclidean length of the curve C, which is defined since we move in a definite C.N.

([2]) The words "Hilbert curve," as used here, correspond to the words "minimizing arc" as used in the introduction.

Let now $U'(P) \subset U(P)$ be chosen in such a way that both the Euclidean and the F-distance of any point Q of $U'(P)$ from the boundary of $U(P)$ are greater than the respective distance of Q from any other point in $U'(P)$. Then a Hilbert arc and the Euclidean segment e_{RS} connecting two points R, S of $U'(P)$ will be contained in $U(P)$, hence we get from (6)

$$AE(e_{RS}) \geqq d(R, S) \geqq BE(e_{RS}); \tag{7}$$

and whenever we have two sequences of curves C_{RS}^m and K_{RS}^m with $E(C_{RS}^m) \to E(e_{RS})$ and $I(K_{RS}^m) \to d(R, S)$ the curves C_{RS}^m and K_{RS}^m will be contained in $U(P)$ from a certain subscript m on. It also follows from (7) that curves with finite generalized length are rectifiable and conversely.

2. **The indicatrix.** We shall now study the function $F(x, \xi) = f(\xi)$ at a fixed point P with coordinates x in a fixed K.N. P will then be the origin of the space of the contravariant vectors ξ. The hypersurface $f(\xi) = 1$, the so-called *indicatrix* $\gamma(x)$ of $F(x, \xi)$ at x, determines $f(\xi)$ on account of its homogeneity. By

$$\frac{\xi^i}{|\xi|} \to \frac{\xi^i}{f(\xi)}, \qquad i = 1, \cdots, n, \tag{1}$$

we map the indicatrix topologically onto the unit sphere $|\xi| = 1$. The inequality (1.2) gives us

$$A|\xi| \geqq f(\xi) \geqq B|\xi|, \tag{2}$$

and for the points ξ on the indicatrix $\gamma(x)$

$$1/B \geqq |\xi| \geqq 1/A, \tag{2'}$$

so that $\gamma(x)$ is between the spheres of radii $1/A$ and $1/B$ around the origin x.

Let a_i be a covariant vector with origin $P = x$. Then $a_i\xi^i = c$ represents a hyperplane in ξ-space. For fixed a_i the function $\phi(\xi) = a_i\xi^i$ is continuous on $\gamma(x)$ and has a positive maximum. For choosing $\xi^i = \rho a_i$ with a suitable $\rho > 0$ we get from (2')

$$\max \phi(\xi) \geqq |a|/A. \tag{3}$$

If $^*\xi$ and $^{**}\xi$ are two points at which this maximum is reached, the equations

$$a_i(\xi^i - {}^*\xi^i) = 0, \qquad a_i(\xi^i - {}^{**}\xi^i) = 0$$

represent the same plane π, which is called a *supporting plane* of $\gamma(x)$ (at each of the points of $\gamma \cdot \pi$). For each $^*\xi$ in $\gamma\pi$ we have, on account of (3),

$$a_i{}^*\xi^i \geqq |a|/A. \tag{3'}$$

For an arbitrary point ξ of $\gamma(x)$ we have

$$a_i\xi^i \leqq a_i{}^*\xi^i,$$

and therefore $a_i(\xi^i - {}^*\xi^i) \leqq 0$, where the equality only holds for $\xi \subset \gamma \cdot \pi$. If π is any plane not containing the origin, we call π' the closed half-space bounded by π which contains it. Then $\gamma(x)$ and also the domain $\Gamma: f(\xi) \leqq 1$ bounded by γ are completely contained in the half-space π'. This shows that our definition of a supporting plane coincides with the usual one. We have

If π: $a_i(\xi^i - {}^*\xi^i) = 0$, ${}^*\xi \subset \gamma \cdot \pi$, is a supporting plane of γ, then

$$f(\eta) - f({}^*\xi) \geqq 0 \qquad \text{if } a_i(\eta^i - {}^*\xi^i) \geqq 0; \tag{4}$$

and conversely, if $f({}^*\xi) = 1$ and if (4) holds, $a_i(\xi^i - {}^*\xi^i) = 0$ will be a supporting plane of γ (or Γ) at ${}^*\xi$.

If we replace a_i by ca_i, $c > 0$, the supporting plane π will not change. (3′) shows that we can find a $\kappa > 0$ such that with $A_i = \kappa a_i$

$$A_i{}^*\xi^i = f({}^*\xi) \tag{5}$$

for every vector ${}^*\xi$ in $\gamma\pi$. (5) will hold for any vector $c\xi^i$, $c > 0$, if it holds for ξ.

We call A_i the *normalized normal* of π and every vector $\xi \neq 0$ which satisfies (5) a *supporting vector* belonging to A_i. Among these supporting vectors those with $f({}^*\xi) = 1$ are called *normalized.*

We can now say

THEOREM 1. *If A_i is a normalized normal of π, ξ^i a fixed supporting vector belonging to A_i, then*

$$f(\eta) - f(\xi) - A_i(\eta^i - \xi^i) \geqq 0, \tag{6}$$

and the equality sign holds if, and only if, $A_i(\eta^i - \xi^i) = 0$.

Remark. Replacing $f(\eta)$ and $f(\xi)$ by $F(x, \eta)$ and $F(x, \xi)$ one recognizes in the left side of (6) a generalization of the Weierstrass $\mathcal{E}$-function.

Proof. With the help of (5) we reduce (6) to

$$f(\eta) - A_i\eta^i \geqq 0. \tag{6'}$$

The left side being homogeneous in the η^i, (6′) will be proved if it is true for all η^i for which $f(\eta) = f(\xi)$; hence it is sufficient to prove (6) for $f(\eta) = f(\xi)$. But then (6) reduces to $-A_i(\eta^i - \xi^i) \geqq 0$ or

$$A_i\left(\frac{\eta^i}{f(\eta)} - \frac{\xi^i}{f(\xi)}\right) \leqq 0;$$

and now $\eta^i/f(\eta)$ is an arbitrary vector on γ, π a supporting plane at $\xi^i/f(\xi)$, and the theorem follows from our previous results.

The surface $f(\xi) = 1$ will be (strictly) convex, if the set $f(\xi) \leqq 1$ is (strictly) convex, and conversely. We quote the following known theorem (compare [2, pp. 21, 22]).

The indicatrix $f(\xi) = 1$ is convex if, and only if,

$$(7) \qquad f(\xi + \eta) \leqq f(\xi) + f(\eta)$$

for arbitrary contravariant vectors ξ^i and η^i. The convexity will be strict if, and only if, in (7) the equality sign holds only for

$$\lambda\xi^i = \mu\eta^i, \qquad \lambda \geqq 0, \qquad \mu \geqq 0, \qquad \lambda + \mu > 0.$$

If the indicatrix of F at x is (strictly) convex in one coordinate system, it will be (strictly) convex in every coordinate system.

If γ is not convex we consider its convex closure $\bar{\Gamma}$ which is the same as the convex closure of Γ. By $\bar{\gamma}(x)$ we designate the boundary of $\bar{\Gamma}$. We shall make use of the following fact (cf. [2, p. 6]).

Every supporting plane of $\gamma(x)$ is a supporting plane of $\bar{\gamma}(x)$ and (which is the less trivial statement) every supporting plane of $\bar{\gamma}(x)$ is a supporting plane of $\gamma(x)$.

Let now $\eta^i \neq 0$ be any vector with $f(\eta) = 1$ and denote by $r(\eta)$ the ray carrying η. Consider the intersections $\beta\eta^i$, $\beta > 0$, of the different supporting planes of γ with $r(\eta)$. We have $\beta \geqq 1$. Let $\alpha(\eta)\eta^i$ be the point where $\bar{\gamma}(x)$ hits $r(\eta)$ and define

$$(8a) \qquad \alpha(x, c\eta) = \alpha(\eta) \qquad \text{for } c > 0,$$

then

$$(8b) \qquad \alpha(x, \xi) \geqq 1 \qquad \text{for all } x \text{ and } \xi$$

and we have

$$(8') \qquad \alpha(\eta) = \min \beta.$$

For since a supporting plane of $\bar{\gamma}$ at $\alpha(\eta)\eta^i$ is a supporting plane of γ, $\alpha(\eta)$ belongs to the numbers β. There can be no $\beta < \alpha(\eta)$ since the corresponding supporting plane of $\gamma(x)$ would separate $\alpha(\eta)\eta^i$ from the origin x (that is $\xi = 0$), and could therefore be no supporting plane of $\bar{\gamma}(x)$.

Since γ is bounded and closed, and $\alpha(\eta)\eta^i$ belongs to the convex closure of γ, we can find points $\xi_1^i, \cdots, \xi_n^i$ on γ such that $\alpha(\eta)\eta^i$ belongs to the simplex with vertices $\xi_1^i, \cdots, \xi_n^i$ (cf. [2, p. 9]). For later use we formulate this statement as follows:

THEOREM 2. *Given a contravariant vector* $(x, \eta) \neq (x, 0)$ *one can find contravariant vectors* $(x, \xi_1), \cdots, (x, \xi_n)$ *with*

$$F(x, \eta) = F(x, \xi_1) = \cdots = F(x, \xi_n)$$

such that

$$(9) \qquad \alpha(x, \eta) \cdot (x, \eta) = \sum_{\alpha=1}^{n} p_\alpha \cdot (x, \xi_\alpha), \qquad p_\alpha \geqq 0, \quad \sum_{\alpha=1}^{n} p_\alpha = 1.$$

We now define a new function $\bar{F}(x, \xi)$ by

$$(10) \qquad \bar{F}(x, \xi) = \bar{f}(\xi) = \frac{1}{\alpha(x, \xi)} F(x, \xi) = \frac{1}{\alpha(\xi)} f(\xi).$$

$\bar{f}(\xi)$ like $f(\xi)$ is positive definite and positive homogeneous of order 1 in ξ.

We see from (8b) that

$$(10') \qquad \bar{F}(x, \xi) \leqq F(x, \xi).$$

$\bar{F}(\alpha, \xi)=1$ is the equation of $\bar{\gamma}=\bar{\gamma}(x)$, since $\alpha(\xi)\xi^i$ is on $\bar{\gamma}$ with ξ^i on γ. With γ the surface $\bar{\gamma}$ is between the spheres of radii $1/A$ and $1/B$, i.e., we still have

$$A\,|\,\xi\,| \geqq \bar{f}(\xi) \geqq B\,|\,\xi\,|,$$

and for $\xi \subset \bar{\gamma}$

$$1/B \geqq |\,\xi\,| \geqq 1/A.$$

The indicatrix γ will be convex if, and only if, $\bar{f}(\eta) \equiv f(\eta)$ or $\alpha(\eta) \equiv 1$. We prove

THEOREM 3. *$\bar{F}(x, \eta)$ is a continuous function of the $2n$ variables $x_1, \cdots, x_n, \eta', \cdots, \eta^n$. Therefore $\alpha(x, \eta)$ is also continuous.*

Proof. Let $\eta_m \rightarrow \eta_0 \neq 0$, $x_m \rightarrow x_0$. Let P_m be the point at which the ray carrying η_m intersects $\bar{\gamma}(x_m)$ (P_m is the point $\alpha(x_m, \eta_m)\cdot(\eta^i_m/F(x_m, \eta_m))$ in the vector space with origin x_m). Let π_m be a supporting plane of $\bar{\gamma}(x_m)$ at P_m. π_m is also a supporting plane of $\gamma(x_m)$ at some point Q_m. We may assume that Q_m converges toward a point Q and π_m toward a plane π at x_0. (Hereby we make use of the fact that a sequence will converge toward P, if every subsequence contains a subsequence which converges to P.) Then P_m will converge to the intersection P of π with the ray $r(\eta_0)$ issuing from x_0. We have to prove that P is the point where $r(\eta_0)$ intersects $\bar{\gamma}(x_0)$, i.e., the point $\alpha(x_0, \eta_0)(\eta_i/F(x_0, \eta_0))$. Since $F(x, \eta)$ is continuous Q will be on $\gamma(x_0)$. Since π passes through Q it is a supporting plane of $\gamma(x_0)$ at Q. Otherwise π would separate x_0 from a point $\bar{\eta}^i$ of $\gamma(x_0)$ and also from all points of a certain closed sphere σ: $\sum(\xi^i-\bar{\eta}^i)^2 \leqq \rho$. For large m the plane π_m would also separate the sphere from the origin x_m, but σ would contain points of $\gamma(x_m)$.

Since π is a supporting plane of $\gamma(x_0)$ and passes through P we see from (8′) that P is either outside or on $\bar{\gamma}(x_0)$. If P was outside of $\bar{\gamma}(x_0)$ we could find a plane π^* separating a certain closed sphere σ^* around P from $\bar{\gamma}(x_0)$ (see [2, p. 4]). For large m, the plane π^* would also separate σ^* from $\gamma(x_m)$ and therefore P_m from $\gamma(x_m)$, hence P_m could not belong to the convex closure of $\gamma(x_m)$, but it lies on $\bar{\gamma}(x_m)$.

Remark on Theorem 3. The proof contains the well known fact:

If $\gamma(x_m)$ is convex and $\gamma(x_m)\rightarrow\gamma(x_0)$ and if π_m is a supporting plane of $\gamma(x_m)$, then each accumulation plane of the π_m is a supporting plane of $\gamma(x_0)$.

If the indicatrix $\gamma(x)$ is strictly convex, a supporting plane $\sum a_i\eta^i=c>0$

can have only one common point with $\gamma(x)$. Hence, given a covariant vector $a_i \neq 0$ there is only one normalized supporting vector ξ^i belonging to α_i; ξ^i is a single-valued function of a_i with

$$\xi^i(\lambda a_i) = \xi^i(a) \qquad \text{for } \lambda > 0.$$

If, conversely, ξ^i is a point on $\gamma(x)$, a supporting plane $\sum a_i \eta^i = c > 0$ through ξ exists; hence ξ belongs to a, but a will in general not be uniquely determined.

We call $F(x, \xi)$ *quasi-regular at x in the direction* ξ, or ξ a quasi-regular direction of F at x if $\alpha(x, \xi) = 1$. If ξ is quasi-regular, supporting planes of $\gamma(x)$ through ξ will exist. If each such supporting plane meets γ only at ξ, the *direction ξ is called regular*. If all directions of $F(x, \xi)$ at x are quasi-regular (regular), *$F(x, \xi)$ is called quasi-regular* (*regular*) *at x*. We have that $F(x, \xi)$ is (regular) quasi-regular at x if, and only if, the indicatrix $F(x, \xi) = 1$ is (strictly) convex.

We shall later make essential use of the

THEOREM 4. *If $F(x, \xi)$ is regular at all points of a K.N., the functions*

$$\xi^i = \xi^i(x, a)$$

are defined and single-valued for $q_i \neq 0$ and continuous in the $2n$ variables $x_1, \cdots, x_n, a_1, \cdots, a_n$.

If F is regular, it follows from the last remark that $F(x, \xi) = 1$ is strictly convex, and therefore $\xi^i(x, e)$ is defined and single-valued for $q_i \neq 0$. Now let $x_m \to x_0$ and $a_m \to a_0$. There is exactly one supporting plane π_m of $F(x_m, \eta) = 1$ of the form $\sum a_{m,i} \eta^i = c_m > 0$. This plane has only the point $\xi_m = \xi(x_m, a_m)$ in common with $\gamma(x_m)$. Each accumulation plane π of the π_m is a supporting plane of $F(x_0, \eta) = 1$ according to the Remark on Theorem 3. Now the normals a_m of π_m converge to a_0; therefore each such π has the form $\sum a_{0,i} \eta^i = c_0 > 0$. But there is only one supporting plane π_0 of $\gamma(x_0)$ of this form; hence the π_m converge to π_0. The accumulation points of the ξ_m must be in $\pi_0 \cdot \gamma(x_0)$, but since $\gamma(x_0)$ is strictly convex $\pi_0 \cdot \gamma(x_0)$ consists of exactly one point ξ_0; therefore, $\xi_m \to \xi_0 = \xi(x_0, a_0)$, which completes the proof.

3. The relation between the F-metric and the $\overline{F}$-metric. The functions $F(x, \xi)$ and $\overline{F}(x, \xi)$ are both continuous as functions of the $2n$ variables in a certain C.N. U and for all ξ. If κ is a compact subset of U and σ the unit sphere $|\xi| = 1$, the product $\kappa \times \sigma$ will be compact; hence F and $\overline{F}$ will be uniformly continuous on $\kappa \times \sigma$. We shall apply this fact in the following form.

Given an $\epsilon > 0$ we can find a constant $\delta > 0$ such that for $x, \bar{x} \subset \kappa$; $\xi, {}^*\xi \subset \sigma$; and $|x - \bar{x}| < \delta$, $|\xi - {}^*\xi| < \delta$ the inequalities

$$(1) \qquad |F(\bar{x}, {}^*\xi) - F(x, \xi)| < \epsilon, \qquad |\overline{F}(\bar{x}, {}^*\xi) - \overline{F}(x, \xi)| < \epsilon$$

hold

If C_{PQ} is a curve of class D' from P to Q, $x(t)$, $a \leqq t \leqq b$, a representation of it, we put

$$\bar{I}(C_{PQ}) = \int_a^b \bar{F}(x, \dot{x})dt \tag{2}$$

and call it the $\bar{F}$-length of C_{PQ}. Since $\bar{F} \leqq F$ we have

$$\bar{I}(C_{PQ}) \leqq I(C_{PQ}). \tag{3}$$

Calling $h(P, Q)$ the greatest lower bound of $\bar{I}(C_{PQ})$ for all curves of class D' leading from P to Q we have

$$h(P, Q) \leqq d(P, Q).$$

The main purpose of this section is to prove that the equality holds. It will be sufficient to see that for a given $\epsilon^* > 0$ and a given curve C_{PQ} of class D' a C^*_{PQ} of the same class exists with

$$| \bar{I}(C_{PQ}) - I(C^*_{PQ}) | < \epsilon^*. \tag{4}$$

C_{PQ} can be decomposed into a finite number of arcs each of which is covered by one C.N.; we may therefore restrict ourselves to proving (4) for each such arc.

Now let $P(t)$, $a \leqq t \leqq b$, be a parametrized representation of C_{PQ} and $\pi_m = (a = t_0 < t_1 < \cdots < t_m = b)$ a subdivision of (a, b) with x_i the coordinates of $P_i = P(t_i)$. We form

$$P_m = \sum_{k=0}^{m-1} F(x_k, \Delta x_k), \qquad \bar{P}_m = \sum_{k=0}^{m-1} \bar{F}(x_k, \Delta x_k) \tag{5}$$

and

$$P'_m = \sum_{k=0}^{m-1} \int_{e_{P_k P_{k+1}}} F(x, \dot{x})dt, \qquad \bar{P}'_m = \sum_{k=1}^{m-1} \int_{e_{P_k P_{k+1}}} \bar{F}(x, \dot{x})dt \tag{5'}$$

where $\Delta x_k = x(t_{k+1}) - x(t_k)$ and e_{RS} designates the oriented Euclidean segment from R to S. It is known (see [1, pp. 286–289]) that for a given $\epsilon_1 > 0$ a $\delta_1 > 0$ can be found such that if $\max_k (t_{k+1} - t_k) < \delta_1$ one has

$$| I(C_{PQ}) - P_m | < \epsilon_1, \qquad | I(C_{PQ}) - P'_m | < \epsilon_1, \tag{6}$$

$$| \bar{I}(C_{PQ}) - \bar{P}_m | < \epsilon_1, \qquad | \bar{I}(C_{PQ}) - \bar{P}_m | < \epsilon_1, \tag{6'}$$

and therefore also

$$| P_m - P'_m | < 2\epsilon_1, \qquad | \bar{P}_m - \bar{P}'_m | < 2\epsilon_1. \tag{7}$$

For $\Delta x_k \neq 0$ we put

$$\eta^i = \frac{\Delta x_{k,i}}{\alpha(x_k, \Delta x_k)} = \frac{\Delta x_{k,i}}{\alpha(x_k, \eta)}. \tag{8}$$

According to Theorem 2.2 we can find vectors $\xi_1, \cdots, \xi_n$ with

$$F(x_k, \eta) = F(x_k, \xi_1) = \cdots = F(x_k, \xi_n) \tag{9}$$

such that

$$\alpha(x_k, \eta)\eta^i = \sum_{\alpha=1}^{n} p_\alpha \xi_\alpha^i \text{ with } p_\alpha \geqq 0 \text{ and } \sum p_\alpha = 1. \tag{9'}$$

We then have (see 2.10)

$$\overline{F}(x_k, \Delta x_k) = \overline{F}(x_k, \alpha(x_k, \eta)\eta) = F(x_k, \eta)$$

and

$$\sum_{\alpha=1}^{n} F(x_k, p_\alpha \xi_\alpha^i) = F(x_k, \eta).$$

Hence

$$\overline{F}(x_k, \Delta x_k) = \sum_{\alpha=1}^{n} F(x_k, p_\alpha \xi_\alpha), \tag{10}$$

where $\Delta x_k = \sum_{\alpha=1}^{n} p_\alpha \xi_\alpha$.

$\sum p_\alpha \xi_\alpha$ may be looked at as a polygon with oriented sides ξ_α. Since x_k is fixed in (10), we have a Minkowskian metric, and the length of this polygon will therefore equal the length of any polygon connecting the origin x_k to the point Δx_k in ξ-space, whose oriented sides are parallel to the ξ_α. In particular, if the length of $e_{P_k P_{k+1}}$ is $\leqq \delta$ we can choose such a polygon λ_k in the sphere of radius δ around P_k. If the subdivision π_m was sufficiently fine we conclude from (3.7) that

$$\left| I(\lambda_k) - \overline{F}(x_k, \Delta x_k) \right| < \epsilon E(\lambda_k),$$

where $E(\lambda_k)$ is the Euclidean length of λ_k. Putting

$$C_{PQ}^* = \sum_{k=0}^{m-1} \lambda_k,$$

we shall have

$$\left| I(C_{PQ}^*) - \overline{P}_m \right| < \epsilon E(C_{PQ}^*). \tag{11}$$

For the sake of simple notation we shall put

$$E(P, Q) = E(e_{PQ}), \qquad I(P, Q) = I(e_{PQ}), \qquad \overline{I}(P, Q) = \overline{I}(e_{PQ}). \tag{12}$$

We then have

$$\frac{E(\lambda_k)}{E(P_k, P_{k+1})} = \frac{\sum p_\alpha |\xi_\alpha|}{|\Delta x_k|} = \frac{\sum p_\alpha |\xi_\alpha|}{\alpha(x_k, \eta)|\eta|}.$$

On account of (9) we may assume that all the vectors $\xi_1, \cdots, \xi_n$ and η are on $F(x_k, \xi) = 1$. It follows from (2.2′) that

$$E(\lambda_k) \leqq \sum p_\alpha \frac{1/B}{1/A} E(P_k, P_{k+1}) = \frac{A}{B} E(P_k, P_{k+1}).$$

Hence

$$E(C^*_{PQ}) \leqq \frac{A}{B} E(C_{PQ}).$$

This, together with (6′) and (11), proves (4) and

THEOREM 1. *For any two points P and Q on $\mathfrak{M}$ one has*

$$h(P, Q) = d(P, Q). \tag{13}$$

Remark. The sums (5) can be formed for any continuous curve C_{PQ}. The limit of P_m and P_m' (∞ admitted) will then exist and be the same for all sequences of subdivisions π_m for which $\max_k (t_{k+1} - t_k) \to 0$. (6) shows that for curves C_{PQ} of class D' this limit equals $I(C_{PQ})$. Hence (5) may be used as a (*generalized*) *definition of the F-length* for continuous curves. This is done in Menger's investigations.

From these considerations we derive

THEOREM 2. *On a minimizing curve M_{PQ} all directions are quasi-regular.*

Proof. Let $P(t)$, $a \leqq t \leqq b$, be a representation of M_{PQ} and assume there is a value t for which $\alpha(x(t), \dot{x}(t)) > 1$. If t is a point of discontinuity for $\dot{x}(t)$ then $\dot{x}_i(t)$ may, for instance, be the right-hand derivatives. Since $\alpha(x, \xi)$ is continuous (Theorem 2.3) we can find an interval $(t, t+\epsilon)$ such that $\alpha(x, \dot{x}) > 1$ in this interval. Then

$$\int_t^{t+\epsilon} \overline{F}(x, \dot{x})dt < \int_t^{t+\epsilon} F(x, \dot{x})dt,$$

and Theorem 3.1 shows that M_{PQ} cannot be minimizing. Theorem 2 can be generalized to relative minimizing curves.

4. **The relation between the F-length and the generalized length.** The following consideration will be local. We may therefore restrict ourselves to a definite K-neighborhood $U(P_0)$ of a given point $P_0 = x_0$. Besides 3.12 we use the notation

$$M(C) = \int_C F(x_0, x')dt, \qquad M(P, Q) = M(e_{PQ}).$$

If the indicatrix of $F(x, \xi)$ is convex at x_0 we have for any curve of class D' from P to Q

$$M(C_{PQ}) \geqq M(P, Q), \tag{1}$$

since M then is simply a Minkowskian metric for which the straight lines are minimizing. If C_ν, $\nu=1, 2, \cdots$, is a sequence of curves of class D' converging to P_0 we have

$$\lim \frac{M(C_\nu)}{I(C_\nu)} = 1; \tag{2}$$

for, on account of the continuity of $F(x, \xi)$, we have for $|\dot{x}|=1$

$$u_\nu = \max_{x<C_\nu} |F(x_0, \dot{x}) - F(x, \dot{x})| \to 0$$

and

$$\left|\frac{M(C_\nu) - I(C_\nu)}{I(C_\nu)}\right| \leqq \frac{\int_{C_\nu} |F(x_0, x') - F(x, \dot{x})| \, dt}{I(C_\nu)} \leqq u_\nu \frac{E(C_\nu)}{B \cdot E(C_\nu)}.$$

Let now $P_\nu \neq Q_\nu$, $P_\nu \to P_0$, $Q_\nu \to P_0$. Then we can find a curve C_ν of D' class from P_ν to Q_ν such that $C_\nu \to P_0$ and

$$0 \leqq I(C_\nu) - d(P_\nu, Q_\nu) < \frac{E(P_\nu, Q_\nu)}{\nu}. \tag{3}$$

Designating by ξ_ν the unit vector in the direction of $e_{P_\nu Q_\nu}$ we get on account of (1)

$$0 \leqq \frac{I(C_\nu)}{M(C_\nu)} - \frac{d(P_\nu, Q_\nu)}{M(C_\nu)} < \frac{E(P_\nu, Q_\nu)}{\nu \cdot M(P_\nu, Q_\nu)} = \frac{1}{\nu \cdot F(x_0, \xi_\nu)},$$

and therefore with the help of (2)

$$\frac{d(P_\nu, Q_\nu)}{M(C_\nu)} \to 1. \tag{4}$$

We have, furthermore, from (1)

$$1 \geqq \frac{d(P_\nu, Q_\nu)}{I(P_\nu, Q_\nu)} = \frac{d(P_\nu, Q_\nu)}{M(C_\nu)} \cdot \frac{M(C_\nu)}{M(P_\nu, Q_\nu)} \cdot \frac{M(P_\nu, Q_\nu)}{I(P_\nu, Q_\nu)} \geqq \frac{d(P_\nu, Q_\nu)}{M(C_\nu)} \cdot \frac{M(P_\nu, Q_\nu)}{I(P_\nu, Q_\nu)}.$$

We see from (2), (4) that the limit of the right side is 1; hence $d(P_\nu, Q_\nu)/I(P_\nu, Q_\nu) \to 1$. Altogether we have the

THEOREM 1. *If the indicatrix of $F(x, \xi)$ at $x_0 = P_0$ is convex (i.e., if F is quasi-regular at x_0), and if $P_\nu \to P_0$, $Q_\nu \to P_0$, $Q_\nu \neq P_\nu$, then*

$$\lim \frac{d(P_\nu, Q_\nu)}{I(P_\nu, Q_\nu)} = \lim \frac{d(P_\nu, Q_\nu)}{M(P_\nu, Q_\nu)} = 1. \tag{5}$$

COROLLARY. *Let $F(x, \xi)$ be quasi-regular at every point of a compact subset σ of a C.N. Then for a given $\epsilon > 0$ a $\delta > 0$ can be found such that*

$$0 \leq \frac{I(P, Q)}{d(P, Q)} - 1 < \epsilon \tag{6}$$

for every pair P, Q in σ with $E(P, Q) < \delta$.

We deduce therefrom

THEOREM 2. *If $F(x, \xi)$ is quasi-regular everywhere on $\mathfrak{M}$, then the F-length coincides with the generalized length for any curve C of class D'.*

$$I(C) = L(C). \tag{7}$$

Therefore a Hilbert arc of class D' will be minimizing.

Proof. Let $P(t)$, $a \leq t \leq b$, be a representation of C and let the subdivision $t_0 = a < t_1 < \cdots < t_n = b$ be chosen in such a way that the arc C_i corresponding to the interval $t_i \leq t \leq t_{i+1}$, $i = 0, \cdots, m-1$, is covered by one coordinate system. We have

$$I(C) = \sum I(C_i), \qquad L(C) = \sum L(C_i).$$

Therefore it is sufficient to prove (7) for every i. The set C_i is compact. On account of the corollary we can therefore find a $\delta > 0$ such that for $t_i \leq \tau' < \tau'' \leq t_{i+1}$ and $\tau'' - \tau' < \delta$ one has

$$0 \leq I(P(\tau'), P(\tau'')) - d(P(\tau'), P(\tau'')) < \epsilon d(P(\tau'), P(\tau'')).$$

Let now $t_i = \tau_0 < \tau_1 < \cdots < \tau_\rho = t_{i+1}$ be a subdivision of (t_i, t_{i+1}) for which $\max_i (\tau_{i+1} - \tau_i) < \delta$. Then

$$\sum_{i=0}^{\rho-1} I(P(\tau_i), P(\tau_{i+1})) - \sum_{i=0}^{\rho-1} d(P(\tau_i), P(\tau_{i+1})) < \epsilon \sum_{i=0}^{\rho-1} d(P(\tau_i), P(\tau_{i+1})) < \epsilon L(C), \tag{8}$$

and the theorem follows from (3.6) and the definition of generalized length.

If we define F-length and generalized length according to the Remark on page 184, formula (8) shows that (7) holds for every continuous curve C.

Another application of Theorem 1 is

THEOREM 3. *If $x_\nu = P_\nu \to P_0 = x_0$ and $\bar{x}_\nu = Q_\nu \to P_0$ in such a way that the univector $\xi_\nu = (\bar{x}_\nu - x_\nu)/|\bar{x}_\nu - x_\nu|$ converges to a vector ξ (i.e., if the oriented straight line $P_\nu Q_\nu$ converges), then*

$$d(P_\nu, Q_\nu)/E(P_\nu, Q_\nu) \to \overline{F}(x_0, \xi). \tag{9}$$

Therefore $\overline{PQ}$ and $\overline{F}(x, \xi)$ determine each other uniquely.

Proof(3). *$\bar{F}(x_0, \xi)=1$ is convex.* We know from (3.13) that $h(P_\nu, Q_\nu) = d(P_\nu, Q_\nu)$; hence we get from (5)

$$\frac{d(P_\nu, Q_\nu)}{\int_{e_{P_\nu Q_\nu}} \bar{F}(x_0, x')dt} = \frac{d(P_\nu, Q_\nu)}{E(P_\nu, Q_\nu)\bar{F}(x_0, \xi_\nu)} \to 1,$$

which proves (9).

As a last application of Theorem 1 we discuss the few differentiability properties of Hilbert arcs which hold under these general conditions.

LEMMA 1. *If $F(x, \xi)$ is quasi-regular at $x_0 = P_0$ and if $P_m \to P_0$, $Q_m \to P_0$, $R_m \to P_0$, and*

$$I_m = \frac{d(P_m, Q_m) + d(Q_m, R_m)}{d(P_m, R_m)} \to 1,$$

then

$$M_m = \frac{M(P_m, Q_m) + M(Q_m, R_m)}{M(P_m, R_m)} \to 1.$$

Proof. $d(Q_m, R_m)/d(P_m, R_m)$ is bounded, since I_m is, and we see from Theorem 1 that

$$\frac{d(Q_m, R_m)}{d(P_m, R_m)}\left[\frac{M(P_m, R_m)}{d(P_m, R_m)} - \frac{M(Q_m, R_m)}{d(Q_m, R_m)}\right] \to 0.$$

Hence we have

$$\lim M_m = \lim \frac{d(P_m, Q_m)M(P_m, Q_m)/d(P_m, Q_m) + d(Q_m, R_m)M(Q_m, R_m)/d(Q_m, R_m)}{d(P_m, R_m)M(P_m, R_m)/d(P_m, R_m)}$$

$$= \lim I_m = 1,$$

We shall prove now

LEMMA 2. *If under the same conditions $F(x, \xi)$ is regular at x_0 we also have*

$$E_m = \frac{E(P_m, Q_m) + E(Q_m, R_m)}{E(P_m, R_m)} \to 1. \tag{10}$$

Proof. By homothetic transformations for each m which may carry x into $\bar{x}$ we can reach

$$\bar{P}_m = P_0, \qquad E(\bar{P}_m, \bar{R}_m) = E(P_0, \bar{R}_m) = 1.$$

We have

(3) This and the following proof follow closely the paper [3].

$$\bar{I}_m = \frac{M(P_0, \bar{Q}_m) + M(\bar{Q}_m, \bar{R}_m)}{M(P_0, \bar{R}_m)} = I_m$$

and similarly $\bar{E}_m = E_m$. If (10) did not hold we could find a sequence of subscripts $n_1, n_2, \cdots$ such that the points $\bar{Q}_{n_i}$ and $\bar{R}_{n_i}$ converge, to $\bar{Q}$ and $\bar{R}$ say, and such that $\bar{E}_{n_i} \to 1+\eta$, $\eta > 0$. Then $\bar{Q}$ cannot lie on the segment $e_{P_0\bar{R}}$. On the other hand we should have

$$\frac{M(P_0, \bar{Q}) + M(\bar{Q}, \bar{R})}{M(P_0, \bar{R})} = 1$$

(on account of Lemma 1). But $F(x_0, \xi)$ being strictly convex this is impossible, because the Euclidean segments are the only minimizing arcs for the M-metric.

We can prove now

THEOREM 4. *Let C_{AB} be a Hilbert arc from A to B and $P_0 = x_0$ an interior point of C_{AB} at which $F(x, \xi)$ is regular. If C_{AB} has at P_0 a left- (right-) hand tangent, it will also have a right- (left-) hand tangent and the two tangents coincide. Hence, if C_{AB} is of class D', it is of class C'.*

Proof. Since C_{AB} is a Hilbert arc it can be mapped topologically onto an interval $a \leqq t \leqq b$. Let $P(t)$ be such a representation, and assume that P_0 corresponds to t_0 and that a left-hand tangent exists. Let $t_\nu \to t_0 + 0$. For every t_ν which is sufficiently close to t_0 we can find a value $\tau_\nu < t_0$ such that $E(P(\tau_\nu), P_0) = E(P_0, P(t_\nu))$. Since C_{AB} is a Hilbert arc, we have

$$\frac{d(P(\tau_\nu), P_0) + d(P_0, P(t_\nu))}{d(P(\tau_\nu), P(t_\nu))} = 1$$

and Lemma 2 tells us that

$$\frac{E(P(\tau_\nu), P_0) + E(P_0, P(t_\nu))}{E(P(\tau_\nu), P(t_\nu))} \to 1. \tag{11}$$

The straight line $P(\tau_\nu)P_0$ converges to the left-hand tangent of C_{AB} at P_0. Since the triangle $P(\tau_\nu)P_0P(\tau_\nu)$ is isosceles it follows from (11) that $\sphericalangle P(\tau_\nu)P_0P(t_\nu)$ tends to π, which proves the theorem.

In spite of this theorem, examples have been given which show that even if $F(x, \xi)$ is regular everywhere, and if all partial derivatives of F with respect to the ξ^i exist, *points A, B may occur for which no Hilbert arc from A to B of class D' exists.* (See [3], pp. 281–285.)

5. **The Lipschitz condition; pseudo-circles.** Consider the last statement together with the example of the Minkowskian geometry, where F does not depend on x. In the latter case no conditions regarding the dependence of F on the ξ^i are put besides the regularity, nevertheless the minimizing arcs are

analytic curves, namely straight lines. One is led to impose on F conditions regarding its variation with respect to the x. The weakest condition which suggests itself is the following *Lipschitz condition*:

To a given point $\bar{x}_0=\bar{P}_0$ there exists a C.N. $U(\bar{P}_0)$ such that for any two points x and $\bar{x}$ in $U(\bar{P}_0)$

(1e) $|F(\bar{x},\xi)-F(x,\xi)|\leqq C|\bar{x}-x|\cdot|\xi|$,

where C depends on $U(\bar{P}_0)$ but not on x, $\bar{x}$ or ξ. Furthermore, we assume

(1f) The function $F(x,\xi)$ is regular everywhere.

From (1e) we derive a simple inequality which is fundamental for the following considerations:

LEMMA 1. *Let $e_{x_0x_1}$ be in $\bar{U}(\bar{P}_0)$ where $\bar{U}(\bar{P}_0)$ is a K.N. contained in the $U(\bar{P}_0)$ of the preceding definition, and let x be an interior point of $e_{x_0x_1}$. Let (x,ξ) and (x_1,ξ_1) be two parallel vectors with*

$$\xi^i|x_0-x_1|=\xi_1^i|x_0-x|. \tag{2}$$

Then

$$F(x_1,\xi_1)-F(x,\xi)\geqq\frac{|\xi||x-x_1|}{|x-x_0|}\{B-C|x-x_0|\}, \tag{3}$$

where B has the same signification as in Lemma 1.1.

Proof. We have

$$F(x_1,\xi_1)-F(x,\xi)=\{F(x_1,\xi_1)-F(x_1,\xi)\}+\{F(x_1,\xi)-F(x,\xi)\}.$$

We see from (2) that

$$\begin{aligned}F(x_1,\xi_1)-F(x_1,\xi)&=F(x_1,\xi)\left[\frac{|x_0-x_1|}{|x_0-x|}-1\right]\\&=F(x_1,\xi)\frac{|x-x_1|}{|x-x_0|}\geqq\frac{B|\xi||x-x_1|}{|x-x_0|}.\end{aligned}$$

We draw from (1e) that

$$|F(x_1,\xi)-F(x,\xi)|<C|x-x_1|\cdot|\xi|.$$

These two equalities prove (3).

From now on we assume that $\mathfrak{M}$ is two-dimensional, $n=2$.

All our considerations will be restricted to a fixed convex K.N. $\bar{U}(\bar{P}_0)$. Then A, B, C may be chosen once for all as positive constants. $k_\rho(P)$ may designate the circle of radius ρ around P and $K_\rho(P)$ the closed circular disk bounded by $k_\rho(P)$.

Let now a point P and two constants ρ and P be chosen in such a way that

$$(4) \qquad \rho < \mathrm{P}, \quad K_{\mathrm{P}}(P) \subset \overline{U}(\overline{P}_0).$$

Let $O = x_0$ be on $k_{\mathrm{P}}(P)$. Then $K_\rho(P)$ does not contain O. There are two continuous vector fields $(x, a) = (x, a; O)$ in $K_\rho(P)$ such that the unit vector a_i is normal to $x - x_0$ at x. Since $F(x, \xi)$ is regular everywhere, we conclude from Theorem 2.4: To $(x, a; O)$ as covariant vector there belongs exactly one vector $(x, \xi) = (x, \xi; O)$ with $F(x, \xi) = 1$ and such that ξ is supporting vector of $\gamma(x)$ to $(x, a; O)$ as normal vector, and $(x, \xi; O)$ will form a continuous vector field in $K_\rho(P)$. As O traverses $k_{\mathrm{P}}(P)$ the vector field $(x, a; O)$ will vary continuously; therefore $(x, \xi; O)$ will depend continuously on x and O. Using the uniform continuity we may say

LEMMA 2. *Given $^*\epsilon > 0$ we can find $\rho > 0$ such that*

$$(5) \qquad |(x, \xi; O) - (\bar{x}, {}^*\xi; O)| < {}^*\epsilon$$

for any two points x, $\bar{x}$ in $K_\rho(P)$ and for every O on $k_{\mathrm{P}}(P)$.

Since $1/B > |\xi| > 1/A$ the inequality (5) implies for the angle ω between ξ and $^*\xi$ that

$$(5') \qquad \omega < \arcsin ({}^*\epsilon A).$$

We designate $K_\rho(P)$ by $K_\rho^*(P)$ if $\arcsin ({}^*\epsilon A) < \frac{1}{2}\epsilon$.

We are now going to define the pseudo-circles. The basic idea of this whole last part will be that these curves have the convexity property which is expressed in Theorem 1 of this section. We keep our previous notations and fix a point O. Consider the differential equations

$$(6) \qquad \frac{dx_i}{dt} = \xi^i(x) = (x, \xi; O), \qquad i = 1, 2.$$

The solutions of (6) are called pseudo-circles with center O. On account of $F(x, \xi) = 1$ the parameter t is the F-length. There passes a solution of (6) through every interior point R of $K_\rho(P)$ which can be continued to either side until it hits $k_\rho(P)$ the first time, at A and B respectively. We see

LEMMA 3. *Every interior point R of $K_\rho(P)$ is contained in at least one pseudo-circular arc σ_{AB} from A to B of class C' and such that A and B are on $k_\rho(P)$ and all other points of σ_{AB} are in the interior of $K_\rho(P)$.*

If $K_\rho(P)$ is a $K_\rho^*(P)$ in the sense defined above, the directions of the tangents of any two pseudo-circles with the same center O (at arbitrary points of the pseudo-circles) differ by less than $\frac{1}{2}\epsilon$.

LEMMA 4. *A Euclidean ray r issuing from O intersects an arc σ_{AB} in at most one point. σ_{AB} is therefore a simple arc.*

For r is never tangential to σ_{AB} as we see from the definition and from

(2.3′). If r had more than one common point with σ_{AB} there would be a subarc σ_{CD} of σ_{AB} whose endpoints are on r and which has no other common points with r. As x traverses σ_{CD} the angle $\sphericalangle COx$ must reach a maximum, for $x=x'$ say, and the ray from O through x' would be tangential to σ_{AB} at x'. It follows that σ_{AB} has a single-valued representation $r=f(\phi)$ in polar coordinates with center O which, as is shown later, is of class C'. Hence the arc σ_{AB} is simple and decomposes $K_\rho(P)$ into exactly two domains σ'_{AB} and σ''_{AB} where σ''_{AB} contains the points (r, ϕ) of $K_\rho(P)$ with $r<f(\phi)$ and is called the convex side of σ_{AB} and σ'_{AB} contains the points (r, ϕ) with $r>f(\phi)$ and is called the concave side of σ_{AB}.

LEMMA 5. *Let R be a given interior point of $K_\rho(P)$ and ξ a given direction through R. There is always a pseudo-circular arc σ_{AB} through R with tangent ξ at $R=x_r$ and such that σ'_{AB} is on the left- (right-) side of σ_{AB} with respect to the orientation of σ_{AB}.*

Proof. Let $F(x_r, \xi)=1$. Let t be a supporting line of $F(x_r, \xi)=1$ at ξ and a_i the unit vector normal to t. Then (x_r, a) will be a supporting normal to ξ_i as supporting vector. Take the straight line perpendicular to a_i through R and let O, O' be its intersections with $k_P(P)$. The pseudo-circles through R with centers O and O' have a tangent with direction ξ at R and the corresponding concave domain is to the right or left of the pseudo-circle according to whether O or O' was taken.

We now come to the main property of the pseudo-circles.

THEOREM 1. *Let $E(O, P)=\mathrm{P}$, $\rho<\mathrm{P}$, $K_{\mathrm{P}}(P)\subset\bar{U}(P_0)$ and $\rho+\mathrm{P}<B/2C$. Let σ_{AB} be a pseudo-circular arc decomposing $K_\rho(P)$ into the concave side σ'_{AB} and the convex side σ''_{AB}. Let R and S be interior points of σ_{AB} and such that R precedes S on σ_{AB}. If σ_{RS} designates the subarc from R to S of σ_{AB} and C_{RS} is any continuous curve from R to S which, except for R and S is completely in the interior of σ'_{AB} then*

$$L(C_{RS}) > L(\sigma_{RS}) = I(\sigma_{RS}). \tag{7}$$

Proof. Since σ_{RS} is rectifiable, we may assume that C_{RS} is rectifiable and also that it is a simple arc. Let $P(t)$, $a\leqq t\leqq b$, be a representation of C_{RS}. Introduce again polar coordinates (r, ϕ) with center $O=x_O$, let $P(x)=(r(t), \omega(t))$, and let σ_{AB} again have the representation $r=f(\phi)$, in particular $(f(\phi_1), \phi_1)=R$, $(f(\phi_2), \phi_2)=S$. We may assume that $\phi_1<\phi_2$. The function $\omega(t)$ is continuous; therefore there is a last value t_1 in (a, b) for which $\omega(t_1)=\phi_1$ and a first value t_2 in (t_1, b) such that $\omega(t_2)=\phi_2$ ($t_1=a$ or $t_2=b$ admitted).

We then have

$$\phi_1 < \omega(t) < \phi_2 \qquad \text{for } t_1 < t < t_2. \tag{8}$$

Put $P(t_1)=R_1$, $P(t_2)=S_1$ and let $C_{R_1S_1}$ be the subarc $t_1\leqq t\leqq t_2$ of C_{RS}. It is

sufficient to prove that

$$I(C_{R_1S_1}) > L(\sigma_{RS}). \tag{9}$$

It follows from (8) that for $t_1 \leqq t \leqq t_2$ the segment $e_{OP(t)}$ will intersect σ_{AB} in exactly one point, namely ${}_*P(t) = (f(\omega(t)), \omega(t))$ which we call the projection of $P(t)$ onto σ_{AB}. We have ${}_*P(t_1) = R$, ${}_*P(t_2) = S$. Let $t' - t_1 > 0$ and $t_2 - t'' > 0$ be so small that ${}_*P(t')$ precedes ${}_*P(t'')$ on σ_{AB} and put $P(t') = R'$, $P(t'') = S'$. For $t' \leqq t \leqq t''$ the number $E(P(t), {}_*P(t))$ is a positive and continuous function of t, which therefore has a positive maximum α. Let now

$$0 < 3\epsilon < \frac{\alpha\gamma C}{2} \tag{10}$$

with $\gamma = (\mathrm{P} - \rho)\left|\omega(t') - \omega(t'')\right|$. We choose $t_1 < t_* < t'$ and $t'' < t^* < t_2$ such that with $\overline{R} = P(t_*)$, $\overline{S} = P(t^*)$ we have

$$L(C_{R_1S_1}) - L(C_{\overline{R}\overline{S}}) < \epsilon, \qquad I(\sigma_{RS}) - I(\sigma_{*\overline{R}*\overline{S}}) < \epsilon. \tag{11}$$

We then choose a subdivision $\pi = (\tau_0 = t_* < \tau_1 < \cdots < \tau_m = t^*)$ $(P(\tau_i) = P_i)$ such that t' and t'' occur among the τ_i and so fine that firstly the chords $e_{P_iP_{i+1}}$ are all contained in the interior of σ'_{AB} and secondly

$$\left| L(C_{\overline{R}\overline{S}}) - \sum_{\nu=0}^{m-1} I(P_\nu, P_{\nu+1}) \right| < \epsilon \tag{12}$$

which is possible on account of the Remark on page 184 and thirdly

$$E(P, {}_*P) > \tfrac{1}{2}\alpha \qquad \text{for } P \subset \sum e_{P_\nu P_{\nu+1}}. \tag{13}$$

Let ${}_*P_\nu$ be the projection of P_ν. We call the subarc $\sigma_{*P_\nu\,*P_{\nu+1}}$ of σ_{AB} regular if ${}_*P_\nu$ precedes ${}_*P_{\nu+1}$ on σ_{AB}. The regular arcs $\sigma_{*P_\nu\,*P_{\nu+1}}$ cover $\sigma_{*\overline{R}*\overline{S}}$, hence designating by $\sum'$ the summation over the regular subarcs,

$$\sum{}' I(\sigma_{*P_\nu\,*P_{\nu+1}}) \geqq I(\sigma_{*\overline{R}*\overline{S}}). \tag{14}$$

Let

$${}_*x_\nu(t), \qquad t_\nu \leqq t \leqq t_{\nu+1},$$

be a representation of the regular arc $\sigma_{*P_\nu\,*P_{\nu+1}}$, where ${}_*x_\nu(t)$ is of class C'. To every point ${}_*x(t)$ there belongs exactly one point $x(t)$ of $e_{P_\nu P_{\nu+1}}$ whose projection ${}_*x(t)$ is. In this way $x(t)$ is defined as function of the same parameter t. $x_i(t)$ is also of class C'. (One sees this immediately if one puts $x(t) = (r(t), \phi(t))$, for then

$$\phi(t) = \operatorname{arc\,tg} \frac{{}_*x_2(t) - x_{02}}{{}_*x_1(t) - x_{01}}$$

where $x_0 = 0$.)

We form

$$I(P_\nu, P_{\nu+1}) - I(\sigma_{*P_\nu *P_{\nu+1}}) = \int_{t_\nu}^{t_{\nu+1}} \{F(x(t), \dot{x}(t)) - F({}_*x(t), {}_*\dot{x}(t))\} dt$$

$$(15) \qquad = \int_{t_\nu}^{t_{\nu+1}} \{F(x, \dot{x}) - F({}_*x, \kappa\dot{x})\} dt$$

$$+ \int_{t_\nu}^{t_{\nu+1}} \{F({}_*x, \kappa\dot{x}) - F({}_*x, {}_*x')\} dt,$$

where $\kappa = |{}_*x - x_0| / |x - x_0|$. By construction the vector $\pm \{\kappa [x_i(t+\Delta) - x_i(t)] - [{}_*x_i(t+\Delta) - {}_*x_i(t)]\}$ is either a zero vector or has the direction of the segment $e_{x_0\,{}_*x}(t+\Delta)$. Letting $\Delta \to 0$ we see that $\pm(\kappa\dot{x}_i(t) - {}_*\dot{x}_i(t))$ has the direction of $e_{x_0\,{}_*x(t)}$. Furthermore, it follows from the definition of pseudo-circles that ${}_*\dot{x}(t)$ is a supporting vector of $F({}_*x, \xi) = 1$ belonging to a covariant vector a_i perpendicular to the ray $\overrightarrow{O_*x}$. Let A_i be the normalized normal in the direction a_i. Applying (2.6) we see that

$$F({}_*x, \kappa\dot{x}) - F({}_*x, {}_*\dot{x}) \geqq A_i(\kappa\dot{x}_i - {}_*\dot{x}_i).$$

Hence the second integral on the right side of (15) is not negative. Only for this conclusion we used the defining property of pseudo-circles.

The Lipschitz condition will be used to get an estimate of the first integral. For we get from (3) that

$$F(x, \dot{x}) - F({}_*x, \kappa\dot{x}) \geqq \frac{\kappa |\dot{x}| \cdot |x - {}_*x|}{|{}_*x - x_0|} \{B - C |{}_*x - x_0|\}$$

$$= \frac{|\dot{x}| \cdot |x - {}_*x|}{|x - x_0|} \{B - C |{}_*x - x_0|\}.$$

Now we have $|{}_*x - x_0| < B/(2C)$, $|x - x_0| < B/(2C)$. Therefore

$$I(P_\nu, P_{\nu+1}) - I(\sigma_{*P_\nu *P_{\nu+1}}) \geqq 0$$

for the regular arcs on $C_{\overline{RS}}$ and on account of (13) we have $|x - {}_*x| > \frac{1}{2}\alpha$ for the regular arcs on $C_{R'S'}$. Hence we have for all these arcs, using the above inequalities for $|{}_*x - x_0|$ and $|x - x_0|$,

$$F(x, \dot{x}) - F({}_*x, \kappa\dot{x}) \geqq |\dot{x}| \frac{\alpha C}{2}.$$

We then get from (15) that

$$I(P_\nu, P_{\nu+1}) - I(\sigma_{*P_\nu *P_{\nu+1}}) \geqq \int_{t_\nu}^{t_{\nu+1}} |\dot{x}| \frac{\alpha C}{2} dt = \frac{\alpha C}{2} E(P_\nu, P_{\nu+1}),$$

and therefrom, using (14),

$$
(16)\qquad \begin{aligned} \sum I(P_\nu, P_{\nu+1}) &\geqq \sum{}' I(P_\nu, P_{\nu+1}) \geqq \sum{}' I(\sigma_{\bar{P}_\nu \bar{P}_{\nu+1}}) + \sum{}'' \frac{\alpha C}{2} E(P_\nu, P_{\nu+1}) \\ &\geqq I(\sigma_{\bar{R}\bar{S}}) + \frac{\alpha C}{2} \sum{}'' E(P_\nu, P_{\nu+1}), \end{aligned}
$$

where $\sum''$ means the sum over all regular $e_{P_\nu P_{\nu+1}}$ for which P_ν and $P_{\nu+1}$ belong to $C_{R'S'}$.

To get an estimate for $\sum'' E(P_\nu, P_{\nu+1})$ we remark that the $e_{P_\nu P_{\nu+1}}$ are outside of $K_{\mathrm{P}-\rho}(O)$. Therefore

$$
E(P_\nu, P_{\nu+1}) > |\omega(\tau_{\nu+1}) - \omega(\tau_\nu)| (\mathrm{P} - \rho),
$$

and since the projections of the regular segments $e_{P_\nu P_{\nu+1}}$ cover the arc of $K_{\mathrm{P}-\rho}(0)$ from $\omega(t')$ to $\omega(t'')$ we have

$$
(17)\qquad \sum{}'' E(P_\nu, P_{\nu+1}) > |\omega(t') - \omega(t'')| \cdot (\mathrm{P} - \rho).
$$

Altogether we get from (10), (11), (12), (16) and (17) that

$$
L(C_{R_1S_1}) - I(\sigma_{RS}) > \frac{\alpha\gamma C}{2} - 3\epsilon > 0,
$$

which proves the theorem.

Since $C_{R_1S_1}$ instead of C_{RS} occurs in the last inequality, we can pronounce the following result.

THEOREM 1'. *Let ρ and* P *satisfy the conditions of Theorem* 1. *Let* Σ_{AB} be *a simple oriented arc which, except for its endpoints A, B is in the interior of $K_\rho(P)$. Let Σ_{AB} contain a pseudo-circular subarc $\sigma_{A'B'}$ whose orientation coincides with that of Σ_{AB}. Designate by Σ'_{AB}, Σ''_{AB} the sets into which Σ_{AB} decomposes $K_\rho(P)$, let Σ'_{AB} be the one whose points close to interior points of $\sigma_{A'B'}$ are on the concave side of $\sigma_{A'B'}$ and finally let R and S be two points of $\sigma_{A'B'}$ where R precedes S. Then* (7) *will hold for any continuous arc C_{RS} from R to S which, except for R and S, is in the interior of Σ'_{AB}.*

6. **Proof that every Hilbert arc is of class C'.** In this section we are going to prove that under the conditions (1a)–(1f) every Hilbert arc is of class C'. We shall see first that a Hilbert arc from Q to P has a tangent at P; i.e., we shall prove

THEOREM 1. *A simple continuous arc C_{QP} from Q to P which has no tangent at P is no Hilbert arc.*

Proof. Since C_{QP} has no tangent at P the rays $\overrightarrow{RP}$ must have at least two different accumulation directions $-g_1$ and $-g_2$ as R tends to P on C_{QP}. Let ϕ and ψ be the two angles between $+g_1$ and $+g_2$ so that $\phi+\psi=2\pi$. By $\omega_\eta(h, T)$ we designate generally a half-angle with h as bisector, opening η and vertex T.

Let now $+g$ and $-g$ be the bisectors of ϕ and ψ respectively, and choose η and ϵ such that the 4 angles

(1) $$\omega_\eta(+g_1, P),\ \omega_\eta(+g_2, P),\ \omega_{3\epsilon}(+g_1, P),\ \omega_{3\epsilon}(-g, P)$$

are disjoint except for their common vertex P.

Take now a circle $K_\rho^*(P)$ (cf. p. 190, after (5′)), and let C_{QP} be represented by $P(t)$, $a\leqq t\leqq b$, and let t_ρ be the greatest number for which $E(P(t), P)=\rho$. Call C_ρ the subarc $t_\rho\leqq t\leqq b$ of C_{QP}. There is a subarc $C_{P_1P_2}$ of C_ρ which connects a point P_1 of $\omega_\eta(+g_1)$ to a point P_2 of $\omega_\eta(+g_2)$ and which is completely in either

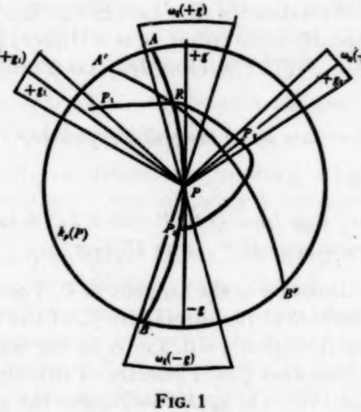


FIG. 1

$\omega_{\phi+\eta}(+g, P)$ or $\omega_{\psi+\eta}(-g, P)$. Assume, for instance, that $C_{P_1P_2}\subset\omega_{\phi+\eta}(+g)$. We then construct a pseudo-circular arc σ_{AB} with center on $k_P(P)$ which passes through P, has $-g$ as tangent at P and for which P_1 is in σ''_{AB} and P_2 in σ'_{AB}. This is possible on account of Lemma 5.4 and because σ_{AB} is contained in $\omega_\epsilon(+g, P)+\omega_\epsilon(-g, P)$ (the tangent of σ_{AB} varies in $K_\rho(P)$ by less than $\frac{1}{2}\epsilon$). Therefore $C_{P_1P_2}$ intersects σ_{AB}. Among the intersections there is one, R say, such that C_{RP_2} belongs to σ'_{AB} except for R. We now draw through R a pseudo-circular arc $\sigma_{B'A'}$ with center O' on $k_\rho(P)$ with the following properties: The tangent of $\sigma_{B'A'}$ at R forms with $+g$ the angle ϵ and such that it points into σ''_{AB}, furthermore such that if σ'_{AB} is on the left (right) of σ_{AB}, the set $\sigma'_{B'A'}$ is on the right (left) of $\sigma_{B'A'}$. The subarc $\sigma_{RA'}$ of $\sigma_{B'A'}$ will lie in σ''_{AB} except for R. Call ${}^*\sigma_{A'R}$ the arc $\sigma_{RA'}$ in opposite orientation. Then the arc $\Sigma_{A'B}={}^*\sigma_{A'R}+\sigma_{RB}$ will be simple. $\omega_\eta(+g_1)$ and $\omega_\eta(+g_2)$ are on different sides of $\Sigma_{A'B}$ since $\Sigma_{A'B}$ lies in $\omega_{3\epsilon}(+g, P)+\omega_{3\epsilon}(-g, P)$. The arc C_{P_2P} contains points of $\omega_\eta(+g_1, P)$; therefore traversing C_{P_2P} we hit $\Sigma_{A'B}$ the first time, at P_3 say. Then C_{RP_3} does not intersect $\Sigma_{A'B}$ except at R and P_3 because ${}^*\sigma_{A'R}$ is contained in σ''_{AB} and C_{RP_3} in σ'_{AB}. If $P_3\subset\sigma_{RA'}$ we see from Theorem 5.1′ that

$$
(16)\qquad \begin{aligned} \sum I(P_\nu, P_{\nu+1}) &\geqq \sum{}' I(P_\nu, P_{\nu+1}) \geqq \sum{}' I(\sigma_{\cdot P_\nu \cdot P_{\nu+1}}) + \sum{}'' \frac{\alpha C}{2} E(P_\nu, P_{\nu+1}) \\ &\geqq I(\sigma_{\cdot \bar{R} \cdot \bar{S}}) + \frac{\alpha C}{2} \sum{}'' E(P_\nu, P_{\nu+1}), \end{aligned}
$$

where $\sum''$ means the sum over all regular $e_{P_\nu P_{\nu+1}}$ for which P_ν and $P_{\nu+1}$ belong to $C_{R'S'}$.

To get an estimate for $\sum'' E(P_\nu, P_{\nu+1})$ we remark that the $e_{P_\nu P_{\nu+1}}$ are outside of $K_{\mathrm{P}-\rho}(O)$. Therefore

$$
E(P_\nu, P_{\nu+1}) > \left| \omega(\tau_{\nu+1}) - \omega(\tau_\nu) \right| (\mathrm{P} - \rho),
$$

and since the projections of the regular segments $e_{P_\nu P_{\nu+1}}$ cover the arc of $K_{\mathrm{P}-\rho}(0)$ from $\omega(t')$ to $\omega(t'')$ we have

$$
(17)\qquad \sum{}'' E(P_\nu, P_{\nu+1}) > \left| \omega(t') - \omega(t'') \right| \cdot (\mathrm{P} - \rho).
$$

Altogether we get from (10), (11), (12), (16) and (17) that

$$
L(C_{R_1 S_1}) - I(\sigma_{RS}) > \frac{\alpha\gamma C}{2} - 3\epsilon > 0,
$$

which proves the theorem.

Since $C_{R_1 S_1}$ instead of C_{RS} occurs in the last inequality, we can pronounce the following result.

THEOREM 1'. *Let ρ and P satisfy the conditions of Theorem* 1. *Let Σ_{AB} be a simple oriented arc which, except for its endpoints A, B is in the interior of $K_\rho(P)$. Let Σ_{AB} contain a pseudo-circular subarc $\sigma_{A'B'}$ whose orientation coincides with that of Σ_{AB}. Designate by Σ'_{AB}, Σ''_{AB} the sets into which Σ_{AB} decomposes $K_\rho(P)$, let Σ'_{AB} be the one whose points close to interior points of $\sigma_{A'B'}$ are on the concave side of $\sigma_{A'B'}$ and finally let R and S be two points of $\sigma_{A'B'}$ where R precedes S. Then* (7) *will hold for any continuous arc C_{RS} from R to S which, except for R and S, is in the interior of Σ'_{AB}.*

6. **Proof that every Hilbert arc is of class** C'. In this section we are going to prove that under the conditions (1a)–(1f) every Hilbert arc is of class C'. We shall see first that a Hilbert arc from Q to P has a tangent at P; i.e., we shall prove

THEOREM 1. *A simple continuous arc C_{QP} from Q to P which has no tangent at P is no Hilbert arc.*

Proof. Since C_{QP} has no tangent at P the rays $\overrightarrow{RP}$ must have at least two different accumulation directions $-g_1$ and $-g_2$ as R tends to P on C_{QP}. Let ϕ and ψ be the two angles between $+g_1$ and $+g_2$ so that $\phi+\psi=2\pi$. By $\omega_\eta(h, T)$ we designate generally a half-angle with h as bisector, opening η and vertex T.

Let now $+g$ and $-g$ be the bisectors of ϕ and ψ respectively, and choose η and ϵ such that the 4 angles

$$(1) \qquad \omega_\eta(+g_1, P),\ \omega_\eta(+g_2, P),\ \omega_{3\epsilon}(+g_1, P),\ \omega_{3\epsilon}(-g, P)$$

are disjoint except for their common vertex P.

Take now a circle $K_\rho^*(P)$ (cf. p. 190, after (5′)), and let C_{QP} be represented by $P(t)$, $a \leqq t \leqq b$, and let t_ρ be the greatest number for which $E(P(t), P) = \rho$. Call C_ρ the subarc $t_\rho \leqq t \leqq b$ of C_{QP}. There is a subarc $C_{P_1P_2}$ of C_ρ which connects a point P_1 of $\omega_\eta(+g_1)$ to a point P_2 of $\omega_\eta(+g_2)$ and which is completely in either

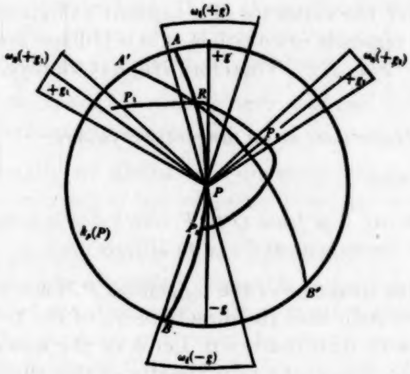


FIG. 1

$\omega_{\phi+\eta}(+g, P)$ or $\omega_{\psi+\eta}(-g, P)$. Assume, for instance, that $C_{P_1P_2} \subset \omega_{\phi+\eta}(+g)$. We then construct a pseudo-circular arc σ_{AB} with center on $k_P(P)$ which passes through P, has $-g$ as tangent at P and for which P_1 is in σ''_{AB} and P_2 in σ'_{AB}. This is possible on account of Lemma 5.4 and because σ_{AB} is contained in $\omega_\epsilon(+g, P) + \omega_\epsilon(-g, P)$ (the tangent of σ_{AB} varies in $K_\rho(P)$ by less than $\frac{1}{2}\epsilon$). Therefore $C_{P_1P_2}$ intersects σ_{AB}. Among the intersections there is one, R say, such that C_{RP_2} belongs to σ'_{AB} except for R. We now draw through R a pseudo-circular arc $\sigma_{B'A'}$ with center O' on $k_\rho(P)$ with the following properties: The tangent of $\sigma_{B'A'}$ at R forms with $+g$ the angle ϵ and such that it points into σ''_{AB}, furthermore such that if σ'_{AB} is on the left (right) of σ_{AB}, the set $\sigma''_{B'A'}$ is on the right (left) of $\sigma_{B'A'}$. The subarc $\sigma_{RA'}$ of $\sigma_{B'A'}$ will lie in σ''_{AB} except for R. Call ${}^*\sigma_{A'R}$ the arc $\sigma_{RA'}$ in opposite orientation. Then the arc $\Sigma_{A'B} = {}^*\sigma_{A'R} + \sigma_{RB}$ will be simple. $\omega_\eta(+g_1)$ and $\omega_\eta(+g_2)$ are on different sides of $\Sigma_{A'B}$ since $\Sigma_{A'B}$ lies in $\omega_{3\epsilon}(+g, P) + \omega_{3\epsilon}(-g, P)$. The arc C_{P_2P} contains points of $\omega_\eta(+g_1, P)$; therefore traversing C_{P_2P} we hit $\Sigma_{A'B}$ the first time, at P_3 say. Then C_{RP_3} does not intersect $\Sigma_{A'B}$ except at R and P_3 because ${}^*\sigma_{A'R}$ is contained in σ''_{AB} and C_{RP_2} in σ'_{AB}. If $P_3 \subset \sigma_{RA'}$ we see from Theorem 5.1′ that

$$d(R, P_3) \leqq I(\sigma_{RP_3}) < L(C_{RP_3}),$$

where σ_{RP_3} is the subarc from R to P_3 of $\sigma_{B'A'}$ and if $P_3 \subset \sigma_{AB}$ we see from the same theorem that

$$d(R, P_3) \leqq I(\sigma_{RP_3}) < L(C_{RP_3}),$$

where σ_{RP_3} is the subarc of σ_{AB} from R to P_3. Therefore in neither case can C_{PQ} be a Hilbert arc, q.e.d.

Let now H_{QP} be a Hilbert arc from Q to P. Then the subarc H_{QR} of H_{QP} is a Hilbert arc for every interior point R of H_{QP}. It therefore has a tangent at R. It then follows from Theorem 4.4 that H_{QP} has a tangent at R. For the point Q we can prove the existence of a tangent as follows.

If H_{PQ}^* is H_{QP} in opposite orientation, it is a Hilbert arc from P to Q with respect to $F^*(x, \xi) = F(x, -\xi)$. Therefore H_{PQ}^* has a tangent at Q; hence H_{QP} has one too. We see

THEOREM 2. *A Hilbert arc has a tangent everywhere.*

We shall prove now

THEOREM 3. *An arc C_{QP} from Q to P which has a tangent everywhere, but whose tangent is not continuous at P, is no Hilbert arc.*

Proof. Let β be the direction of the tangent at P. Then a sequence of points $P_m \to P$ on C_{QP} exists such that the directions γ_m of the tangent of C_{QP} at P_m converge to a line with direction $\gamma \neq \beta$. Let ϕ be the lesser of the two angles formed by β and γ at P and let ξ be a bisector of this angle. Consider $K_\rho^{\phi/2}(P)$ (for definition see page 190). On $k_P(P)$ we choose the point O such that the vector field $(x, \xi; O)$ has the direction ξ at P and such that if σ_{AB} is a pseudo-circular arc of the vector field the direction γ points into σ'_{AB}.

Let $\sigma_{A_mB_m}$ be a pseudo-circular arc with center O through P_m. Then its direction at P_m differs from ξ by less than $\phi/4$, and since ξ bisects $\sphericalangle \beta\gamma$ and $\gamma_m \to \gamma$ and $\overrightarrow{P_mP} = \beta_m \to \beta$ we see that for large m the rays γ_m and β_m will not belong to the angles $\omega_{\phi/2}(-{}^*\xi, P_m)$ or $\omega_{\phi/2}(+{}^*\xi, P_m)$, where ${}^*\xi$ designates the parallel to ξ through P_m. Therefore γ_m will point into the concave side $\sigma'_{A_mB_m}$ of σ_{AB} and β_m into $\sigma''_{A_mB_m}$.

As in the preceding proof we now replace the subarc $\sigma_{A_mP_m}$ of $\sigma_{A_mB_m}$ by another pseudo-circular arc $\sigma_{P_mA'}$ which is a subarc of a pseudo-circular arc $\sigma_{B'A'}$ through P_m with center $O' \subset K_\rho(P)$ where O' is determined as follows: The field vector of $(x, \xi; O')$ at P has direction $-\xi$ and $\sigma'_{B'A'}$ is on the right (left) of $\sigma_{B'A'}$ if $\sigma'_{A_mB_m}$ is on the left (right) of $\sigma_{A_mB_m}$. Then γ will still point into the concave side of any pseudo-circular arc of $K_\rho^{\phi/2}(P)$ with center O'.

Call $\sigma_{A'P_m}^*$ the subarc $\sigma_{P_mA'}$ of $\sigma_{B'A'}$ in opposite orientation. Then

$$\Sigma_{A'B_m} = {}^*\sigma_{A'P_m} + \sigma_{P_mB_m}$$

will be a simple arc, which is contained in the angles $\omega_{\phi/2}({}^*\xi, P_m)$ and

$\omega_{\phi/2}(-{}^*\xi, P_m)$. Therefore γ_m and β_m point into different sides Σ_1 and Σ_2 of $\Sigma_{A'B_m}$.

Let γ_m point into Σ_1.

Now C_{PQ} has tangent γ_m in P_m. Therefore the points on C_{PQ} following P_m and sufficiently close to P_m must be in Σ_1. Since β_m contains P, the point P must be in Σ_2. Therefore, traversing C_{QP} from P_m on we shall meet $\Sigma_{A'B_m}$ a first time, at R say. Then one sees, as in the preceding proof that, no matter whether R belongs to $\sigma_{P_mA'}$ or σ_{P_mB} the arc C_{P_mR} will be longer than the corresponding subarc of $\sigma_{P_mA'}$ or σ_{P_mB}. We see, as in the preceding proof, that H_{QP} has a continuous tangent at Q too, and hence that every Hilbert arc H_{QP} has a continuous tangent everywhere. If we then introduce the arc length s on H_{QP} as parameter, H_{QP} will have a representation $x_i(s)$ where the $x_i(s)$ are of class C'. Therefore H_{QP} is of class C'. We have

THEOREM 4. *If the space $\mathfrak{M}$ is two-dimensional and if the conditions* (1a)–(1f) *hold, every Hilbert arc is of class C'.*

7. **Examples.** Finally we discuss the question whether under the conditions of the last theorem any of the following questions can be answered positively:

(1) *Have the minimizing curves second derivatives*[4]?

(2) *Is there at least one minimizing curve through a given line element?*

(3) *Is here at most one minimizing curve through a given line element?*

The answer to all three questions is negative. In order to construct a simple example we remind the reader of the function $F(x, y; \xi, \eta) = (\xi^2+\eta^2)^{1/2}/y$, which defines a hyperbolic metric in the half-plane $y>0$ of the Euclidean (x, y)-plane. Therefore

$$\frac{(\xi^2+\eta^2)^{1/2}}{y+1}, \qquad \frac{(\xi^2+\eta^2)^{1/2}}{1-y}$$

will define hyperbolic metrics in the half-planes $y>-1$ and $y<1$ respectively.

The integrand

$$F(x, y; \xi, \eta) = \frac{(\xi^2+\eta^2)^{1/2}}{1+|y|}$$

is regular everywhere and we have

$$\begin{aligned} |F(\bar{x}, \bar{y}; \xi, \eta) - F(x, y; \xi, \eta)| &= (\xi^2+\eta^2)^{1/2}\cdot\left|\frac{1}{1+|\bar{y}|} - \frac{1}{1+|y|}\right| \\ &\leqq (|y| - |\bar{y}|)\cdot(\xi^2+\eta^2)^{1/2} \\ &\leqq ((x-\bar{x})^2 + (y-\bar{y})^2)^{1/2}\cdot(\xi^2+\eta^2)^{1/2} \end{aligned}$$

(4) This question has an invariant meaning only if $\mathfrak{M}$ is of class 2, which will be the case in the following examples.

so that the Lipschitz condition (c) is uniformly satisfied. For $y \geqq 0$ ($y \leqq 0$) the minimizing arcs are the semicircles perpendicular to $y=-1$ ($y=1$). Therefore every point on $y=0$ can be joined to $(0, 0)$ by exactly two minimizing arcs. Hence no minimizing curve with $y=0$ as tangent of $(0, 0)$ exists. The same holds for every other point on the x-axis. *This answers* (2) *negatively.*

To get examples for the other two questions we consider

$$F(x, y; \xi, \eta) = \frac{(\xi^2 + \eta^2)^{1/2}}{1 - |y|}.$$

Here the Lipschitz condition (c) holds uniformly for $|y| \leqq r < 1$ therefore in a suitable neighborhood of any given point in $|y| < 1$. For $y > 0$ ($y < 0$) the extremals are now the semicircles perpendicular to $y=1$ ($y=-1$). Since we have at every point $(a, 0)$ of the x-axis one circle of each family tangent to the x-axis at $(a, 0)$, we see that *question* (3) *has a negative answer.*

But this example also settles question (1). Obviously the segment of the x-axis joining $(0, 0)$ to $(a, 0)$ is the only minimizing arc connecting them. Take now a point $(a, 1-1/2^{1/2})$ with $a > 1/2^{1/2}$ and let H be a Hilbert arc connecting $(a, 1-1/2^{1/2})$ to $(0, 0)$. It cannot contain points $(\bar{x}, \bar{y})$ with $\bar{y} < 0$ for then it would contain a point $(b, 0)$ preceding $(\bar{x}, \bar{y})$ and the segment from $(b, 0)$ to $(0, 0)$ would be shorter than the corresponding arc of H.

We know from Theorem 6.4 that H is of class C'. It follows therefrom that H must look as follows: Let C be the semicircle perpendicular to $y=+1$ through $(a, 1-1/2^{1/2})$ which is tangent to the x-axis, at $(a-1/2^{1/2}, 0)$. Then the arc of c from $(a, 1-1/2^{1/2})$ to $(a-1/2^{1/2}, 0)$ together with the segment of the x-axis between $(a-1/2^{1/2}, 0)$ and $(0, 0)$ will constitute H, but *H has no second derivative at* $(a-1/2^{1/2}, 0)$.

The family of minimizing curves issuing from $(0, 0)$ with the x-axis as tangent does not form a field. We were not able to find a counter-example for (3) with the field property.

References

1. O. Bolza, *Vorlesungen über Variationsrechnung*, Leipzig, 1909.

2. T. Bonnesen, and W. Fenchel, *Theorie der konvexen Körper*, Berlin, 1934.

3. H. Busemann, *Lokale Eigenschaften der zu Variationsproblemen gehörigen metrischen Räume*, Fundamenta Mathematicae, vol. 32 (1939), pp. 265–287.

4. H. Busemann, and Willy Feller, *Krümmungseigenschaften konvexer Flächen*, Acta Mathematica, vol. 66 (1935), pp. 1–47.

Illinois Institute of Technology,
Chicago, Ill.
Institute for Advanced Study,
Princeton, N. J.

ON THE GROWTH PROPERTIES OF A FUNCTION OF TWO COMPLEX VARIABLES GIVEN BY ITS POWER SERIES EXPANSION

BY
ABE GELBART

1. **Introduction.** One of the most fundamental formulas in the theory of functions of one complex variable is the Cauchy integral formula. It is of particular value in the Weierstrass-Hadamard approach, i.e., in obtaining properties of a function from the coefficients of its power series expansion. A similar formula cannot be obtained for functions of two complex variables for an arbitrary four-dimensional domain, as is obtained, for instance, for the bicylinder, where the integration is taken over a two-dimensional surface on the boundary. Bergman(1) has shown, however, that for certain domains far more general than those previously considered, i.e., domains bounded by a finite number of analytic hypersurfaces, an analogous formula does exist, the double integral being taken essentially over the two-dimensional surface common to two or more of the analytic bounding hypersurfaces(2).

In this paper we shall obtain growth properties in terms of the coefficients of the power series expansion of a function $f(z_1, z_2)$ of two complex variables analytic in special domains of the type mentioned above; first, with the aid of Bergman's integral formula, along the two-dimensional surfaces common to the bounding hypersurfaces, and then, along a class of two-dimensional surfaces lying in only one of the bounding hypersurfaces and having a line of contact with another bounding hypersurface. We also obtain a mapping theorem which determines from the coefficients a convex region in the f_1f_2-plane, $f(z_1, z_2)=f_1+if_2$, which must be contained in the smallest convex region of the mapping on the f_1f_2-plane of the surfaces considered.

2. **Properties of f associated with $G^2(r)$.** Let us consider a finite four-dimensional domain $\mathfrak{M}^4$ which is bounded by the hypersurfaces

$$(2.1)\qquad \begin{aligned} \overset{3}{s_1}(r) &\equiv E[z_2 = re^{i\lambda_1},\ 0 \leqq \lambda_1 \leqq 2\pi], \\ \overset{3}{s_2}(r) &\equiv E[z_1 = re^{i\lambda_2} + p(\lambda_2)z_2 \equiv h(\lambda_2, z_2),\ 0 \leqq \lambda_2 \leqq 2\pi], \end{aligned}$$

and which depends on a positive parameter r; $p(\lambda_2)$ is assumed merely to have

Presented to the Society in two parts: October 28, 1939, under the title *On functions of two complex variables given by power series expansion*, and December 27, 1939, under the title *On the growth of a function of two complex variables given by its power series expansion on certain hypersurfaces*; received by the editors May 28, 1940.

(1) Bergman [2, 3]. See the bibliography at the end of this paper.

(2) Bergman calls such surfaces "distinguished boundary surfaces."

a first derivative. Let $G^2(r)$ be the two-dimensional surface on the boundary of $\mathfrak{M}^4$ which is the common part of the bounding hypersurfaces, i.e.,

$$G^2(r) \equiv \overset{3}{s_1} \cdot \overset{3}{s_2}. \tag{2.2}$$

THEOREM I. *Given a function $f(z_1, z_2) = \sum_{m,n=0}^{\infty} a_{mn} z_1^m z_2^n$ regular in the domain $\overline{\mathfrak{M}}^4(r)$; if $M(r)$ is the maximum-modulus of $f(z_1, z_2)$ on $G^2(r)$, then*

$$M(r) \geqq \max_{m,n} \frac{r^{m+n} | a_{mn} |}{G(m, n; p)B(p)}, \tag{2.3}$$

where m and n range over all non-negative integral values, $B(p)$ is a constant depending upon p, and $G(m, n; p)$ is a function of m, n, and p, given by

$$\begin{aligned} &1 + \int_0^{n+1} \left(1 + x \frac{1 + \log m}{m}\right)^m \max | p(\lambda_2) |^x dx - \frac{m}{\log p} - \frac{m}{1 + \log m}, && \text{when } \max | p | < 1, m \geqq 1, \\ &1 + \int_0^{n+1} \left(1 + x \frac{1 + \log m}{m}\right)^m \max | p(\lambda_2) |^x dx, && \text{when } \max | p | \geqq 1, m \geqq 1, \\ &\frac{1 - \max | p(\lambda_2) |^{n+1}}{1 - \max | p(\lambda_2) |}, && \text{when } m = 0 \text{ for all } p. \end{aligned} \tag{2.4}$$

Proof of Theorem I. Keeping z_2 constant, say equal to t_2, we obtain for a particular value of z_1, say t_1,

$$f(t_1, t_2) = \frac{1}{2\pi i} \int_0^{2\pi} \frac{f[h(\lambda_2, t_2), t_2][ire^{i\lambda_2} + p'(\lambda_2)t_2]d\lambda_2}{[(re^{i\lambda_2} + p(\lambda_2)t_2) - t_1]}. \tag{2.5}$$

Since the numerator of the integrand is an analytic function of t_2, we again apply the Cauchy integral formula and obtain

$$\begin{aligned} f(t_1, t_2) = \frac{1}{(2\pi i)^2} \int_0^{2\pi} \int_0^{2\pi} & \frac{f[h(\lambda_2, re^{i\lambda_1}), re^{i\lambda_1}]}{[(re^{i\lambda_2} + p(\lambda_2)t_2) - t_1][re^{i\lambda_1} - t_2]} \\ & \cdot [ire^{i\lambda_2} + p're^{i\lambda_1}]ire^{i\lambda_1}d\lambda_1 d\lambda_2. \end{aligned} \tag{2.6}$$

For the mth derivative of $f(t_1, t_2)$ with respect to t_1, we obtain

$$\begin{aligned} \frac{\partial^m f(t_1, t_2)}{\partial t_1^m} = \frac{m!}{(2\pi i)^2} \int_0^{2\pi} \int_0^{2\pi} & \frac{f[h(\lambda_2, re^{i\lambda_1}), re^{i\lambda_1}]}{[(re^{i\lambda_2} + p(\lambda_2)t_2) - t_1]^{m+1}[re^{i\lambda_1} - t_2]} \\ & \cdot [ire^{i\lambda_2} + p'(\lambda_2)re^{i\lambda_1}]ire^{i\lambda_1}d\lambda_1 d\lambda_2. \end{aligned} \tag{2.7}$$

Let

$$H_1 \equiv (re^{i\lambda_2} + p(\lambda_2)t_2) - t_1, \qquad H_2 \equiv re^{i\lambda_1} - t_2.$$

For the nth derivative of $1/H_1^{m+1}H_2$ with respect to t_2, we obtain by Leibnitz' rule

$$\left[1+\sum_{\nu=1}^{n}\frac{(m+\nu)!}{m!\nu!}\left(\frac{H_2}{H_1}p(\lambda_2)\right)^{\nu}\right]\frac{n!}{H^{m+1}H_2^{n+1}}. \tag{2.8}$$

Hence we obtain for $\partial^{m+n}f(t_1, t_2)/\partial t^m\partial t^n$ the expression

$$\begin{aligned}&\frac{m!n!}{(2\pi i)^2}\int_0^{2\pi}\int_0^{2\pi}\frac{f[h,\lambda_1][ire^{i\lambda_2}+p'(\lambda_2)re^{i\lambda_1}]ire^{i\lambda_1}}{[(re^{i\lambda_2}+p(\lambda_2)t_2)-t_1]^{m+1}[re^{i\lambda_1}-t_2]^{n+1}}\\ &\qquad\cdot\left[1+\sum_{\nu=1}^{n}\frac{(m+\nu)!}{m!\nu!}\left(\frac{H_2}{H_1}p\right)^{\nu}\right]d\lambda_1 d\lambda_2.\end{aligned} \tag{2.9}$$

Now

$$a_{mn}=\frac{\partial^{m+n}f(0,0)}{m!n!\partial t_1^m\partial t_2^n}.$$

Hence

$$\begin{aligned}a_{mn}&=\frac{1}{(2\pi i)^2}\int_0^{2\pi}\int_0^{2\pi}\frac{f[h,\lambda_1][ire^{i\lambda_2}+p're^{i\lambda_1}]ire^{i\lambda_1}}{r^{m+n+2}\exp\{i(m\lambda_2+n\lambda_1+\lambda_2+\lambda_1)\}}\\ &\quad\cdot\left[1+\sum_{\nu=1}^{n}\frac{(m+\nu)!}{m!\nu!}(e^{i(\lambda_1-\lambda_2)}p)^{\nu}\right]d\lambda_1 d\lambda_2.\end{aligned} \tag{2.10}$$

Taking the absolute value of a_{mn} we get

$$\begin{aligned}|a_{mn}|&\leqq\frac{1}{4\pi^2}\,\frac{M(r)\max_{0\leqq\lambda_1\leqq 2\pi}[1+|p'(\lambda_2)|]}{r^{m+n}}\\ &\qquad\cdot\left[1+\sum_{\nu=1}^{n}\frac{(m+\nu)!}{m!\nu!}(\max|p|^{\nu})\right]4\pi^2.\end{aligned} \tag{2.11}$$

Now for $m\geqq 1$, it can be shown that

$$\begin{aligned}&1+\sum_{\nu=1}^{n}\left[\frac{(m+\nu)!}{m!\nu!}\max|p(\lambda_2)|^{\nu}\right]\\ &\qquad\leqq 1+\sum_{\nu=1}^{n}\left(1+\nu\,\frac{1+\log m}{m}\right)^{m}|p|^{\nu}.\end{aligned} \tag{2.12}$$

When $|p|<1$ we have

$$\begin{aligned}&1+\sum_{\nu=1}^{n}\frac{(m+\nu)!}{m!\nu!}\max|p(\lambda_2)|^{\nu}\\ &\qquad\leqq 1+\int_1^n\left[1+x\,\frac{1+\log m}{m}\right]^{m}\max|p|^{x}dx-\frac{m}{\log p}-\frac{m}{1+\log m},\end{aligned} \tag{2.13}$$

and when $|p|\geqq 1$,

$$(2.14)\quad 1+\sum_{\nu=1}^{n}\frac{(m+\nu)!}{m!\nu!}\max|p|^{\nu}\leqq\int_{0}^{n+1}\left[1+x\frac{1+\log m}{m}\right]^{m}\max|p|^{x}dx.$$

When $m=0$, $(1-\max|p|^{n+1})/(1-\max|p|)$ is the exact value of the left-hand side of (2.14) for all p.

Therefore for all differentiable $p(\lambda_2)$ and non-negative integral values of m and n we have

$$(2.15)\qquad |a_{mn}|\leqq\frac{M(r)B(p)}{r^{m+n}}G(m,n;p),$$

where $B(p)=\max_{0\leqq\lambda_2\leqq 2\pi}(1+|p'|)$, or

$$(2.16)\qquad M(r)\geqq\frac{r^{m+n}|a_{mn}|}{B(p)G(m,n;p)}.$$

To find those values of m and n, say $\mu(r)$ and $\nu(r)$, for which the right-hand expression in (2.16) is maximum for a given r, we take the logarithm of the expression, letting $-\log|a_{mn}|=g_{mn}$ and employ a generalized Newton polygon method. Then

$$(2.17)\qquad \begin{aligned} &g_{mn}-(m+n)\log r+\log B+\log G(m,n)\\ &\qquad\geqq g_{\mu\nu}-(\mu+\nu)\log r+\log B+\log G(\mu,\nu)=C. \end{aligned}$$

We choose m, n, and g_{mn} as the x-, y-, and z-axes, respectively, and plot the points (m, n, g_{mn}). Then the m and n of the first point which lies in the surface $z=x\log r+y\log r-\log G(x,y)-\log B+k$ as this surface is translated along the z-axis from $-\infty$ by varying k, i.e., until $k=C$, are the μ and ν which give the right-hand side of (2.16) a maximum. If there is more than one point lying on the surface, the one with the smaller m is chosen; if the m's are the same, the one with the smaller n is chosen. μ and ν are obviously functions of r.

We then have

$$(2.18)\qquad M(r)\geqq\frac{r^{\mu+\nu}|a_{\mu\nu}|}{BG(\mu,\nu)}.$$

This gives a lower bound for the growth of $f(z_1, z_2)$ along the hypersurface $g^3\equiv S_{r=r_0}^{r_1}G^2(r)$, where r varies continuously.

3. **The mapping of the surface $G^2(r)$.** Let us introduce the function

$$(3.1)\qquad F(f,\alpha)=e^{e^{-i\alpha}f(z_1,z_2)}=\sum_{r,s=0}^{\infty}A_{rs}z_1^r z_2^s,$$

where $0\leqq\alpha\leqq 2\pi$ and f is defined as in the previous sections. The coefficients $\{A_{rs}\}$ are functions of α and a combination of the a_{mn}'s such that $m\leqq r$ and $n\leqq s$. We define the region $R^2(r)$ as the product of the half-planes

$$f_1 \cos \alpha + f_2 \sin \alpha \leqq Q(\alpha, r), \qquad 0 \leqq \alpha < 2\pi, \tag{3.2}$$

where f_1 and f_2 are cartesian coordinates in the f_1f_2-plane, and

$$Q(\alpha, r) \equiv \log \mid A_{\mu\nu}(a, \alpha) \mid + (\mu + \nu) \log r - \log G(\mu, \nu) - \log B. \tag{3.3}$$

THEOREM II. *Let $f(z_1, z_2) = f_1 + if_2$. Then the smallest convex domain enclosing the mapping of $G^2(r)$ on the f_1f_2-plane contains the closed convex region $R^2(r)$ which depends only on the coefficients of the expansion of $f(z_1, z_2)$ and the surface $G^2(r)$.*

This gives a lower bound, so to speak, of the mapping of $G^2(r)$ on the f_1f_2-plane.

Proof of Theorem II. Let

$$P(r) = \max \left| e^{e^{-i\alpha} f(z_1, z_2)} \right| \tag{3.4}$$

on the surface $G^2(r)$; then from (3.4) and (2.18)

$$\begin{aligned} \log P(r) &= \log \mid \exp \{e^{-i\alpha} f^*(z_1, z_2)\} \mid \\ &= \log \mid \exp \{(f_1^* \cos \alpha + f_2^* \sin \alpha)\} \mid \\ &\quad \cdot \mid \exp \{-i(f_1^* \sin \alpha - f_2^* \cos \alpha)\} \mid \end{aligned} \tag{3.5}$$

$$\begin{aligned} &= f_1^* \cos \alpha + f_2^* \sin \alpha \\ &\geqq \log \mid A_{\mu\nu}(a, \alpha) \mid + (\mu + \nu) \log r - \log G(\mu, \nu) - \log B = Q(\alpha, r), \end{aligned} \tag{3.6}$$

where the $*$ indicates that value of f which gives $|P|$ its maximum, for a given α. Now, for each α, $Q(\alpha, r)$ has a fixed value (depending on r). It is clear from (3.6) that at least one point of the mapping, namely, (f_1^*, f_2^*), will lie in the half-plane

$$f_1 \cos \alpha + f_2 \sin \alpha \geqq Q(\alpha, r). \tag{3.7}$$

The region $R^2(r)$ will therefore be contained in the smallest convex domain containing the mapping of $G^2(r)$ on the f_1f_2-plane. Theorem II is then proved.

It is clear that a similar theorem will hold for any surface for which we have a lower bound for the maximum of the function $f(z_1, z_2)$ on the surface. For example, we can state similar theorems for the surfaces considered in §§4 and 5.

4. **Further properties of the function on other surfaces of the type $G^2(r)$.** Let us consider the finite four-dimensional region $\mathfrak{M}^4(r)$ bounded by the three infinite hypersurfaces:

$$\begin{aligned} \overset{3}{s}_1(r) &\equiv E[z_2 = re^{i\lambda_1}, \ 0 \leqq \lambda_1 \leqq 2\pi], \\ \overset{3}{s}_2(r) &\equiv E[z_1 = re^{i\lambda_2} + C_2 z_2, \ 0 \leqq \lambda_2 \leqq 2\pi], \\ \overset{3}{s}_3(r) &\equiv E[z_1 = re^{i\lambda_3} - C_3 z_2, \ 0 \leqq \lambda_3 \leqq 2\pi], \end{aligned} \tag{4.1}$$

where, as above, r is a parameter and C_2 and C_3 are positive constants less than unity. This restriction on C_2 and C_3 is necessary in order that the hypersurfaces of (3.1) form the boundary of a finite closed domain. Let $G_{ks}^2(r)$ be that part of $s_k^3(r) \cdot s_s^3(r)$ which belongs to the boundary of $\mathfrak{M}^4$. Now let

$$G^2(r) \equiv G_{12}^2(r) + G_{13}^2(r) + G_{23}^2(r). \tag{4.2}$$

Let also $g^3 \equiv \mathrm{S}_{r=r_0}^{r_1} G^2(r)$, and $g_{ks}^3 \equiv \mathrm{S}_{r=r_0}^{r_1} G_{ks}^2(r)$, where r varies continuously and $r_1 < \infty$.

Let $f(z_1, z_2)$, as before, be an analytic function regular in $\overline{\mathfrak{M}}^4$. We now apply Bergman's integral formula(3) for functions of two complex variables which states that at a point (t_1, t_2) in $\mathfrak{M}^4$,

$$\begin{aligned} f(t_1, t_2) &= \frac{1}{2} \sum_{k,s}{}' M_{ks}(t_1, t_2) \\ &= \frac{1}{2(2\pi i)^2} \sum_{k,s}{}' \int\int_{B_{ks}^2} \frac{f(\overset{(1)}{\phi_{ks}}, \overset{(2)}{\phi_{ks}}) B_{ks}(t_1, t_2, \lambda_k, \lambda_s) d\lambda_k d\lambda_s}{\Phi_k(t_1, t_2, \lambda_k) \Phi_s(t_1, t_2, \lambda_s)}, \end{aligned}$$

$$B_{ks}(t_1, t_2, \lambda_k, \lambda_s) = \frac{Z_{ks}(t_1, t_2, \lambda_k, \lambda_s)}{(\phi_{ks}^{(1)} - t_1)(\phi_{ks}^{(2)} - t_2)}, \qquad k \neq s,$$

$$\begin{aligned} Z_{ks}(t_1, t_2, \lambda_k, \lambda_s) = \frac{D(\overset{(1)}{\phi_{ks}}, \overset{(2)}{\phi_{ks}})}{D(\lambda_k, \lambda_s)} &[\Phi_s(t_1, t_2, \lambda_s) \Phi_k(t_1, \phi^{(1)}, \lambda_k) \\ &- \Phi_k(t_1, t_2, \lambda_k) \Phi_s(t_1, \phi^{(2)}, \lambda_s)], \end{aligned}$$

where B_{ks}^2 is the surface range of integration. We have in our case

$$\begin{aligned} \Phi_1 &= z_2 - re^{i\lambda_1}, \\ \Phi_2 &= z_1 - re^{i\lambda_2} - C_2 z_2, \\ \Phi_3 &= z_1 - re^{i\lambda_3} + C_3 z_2; \end{aligned} \tag{4.3}$$

$$\begin{cases} z_1 = \overset{(1)}{\phi_{12}} \equiv re^{i\lambda_2} + C_2 re^{i\lambda_1}, \\ z_2 = \overset{(2)}{\phi_{12}} \equiv re^{i\lambda_1}, \end{cases}$$

$$\begin{cases} z_1 = \overset{(1)}{\phi_{13}} \equiv re^{i\lambda_3} - C_3 re^{i\lambda_1}, \\ z_2 = \overset{(2)}{\phi_{13}} \equiv re^{i\lambda_1}, \end{cases} \tag{4.4}$$

$$\begin{cases} z_1 = \overset{(1)}{\phi_{23}} \equiv \dfrac{1}{C_2+C_3} [C_2 re^{i\lambda_3} + C_3 re^{i\lambda_2}], \\ z_2 = \overset{(2)}{\phi_{23}} \equiv \dfrac{1}{C_2+C_3} [re^{i\lambda_3} - re^{i\lambda_2}]; \end{cases}$$

(3) Bergman [2, p. 97] and [3, p. 861].

and consequently,

$$
\begin{aligned}
(4.5)\qquad f(t_1, t_2) &= M_{12}(t_1, t_2) + M_{13}(t_1, t_2) + M_{23}(t_1, t_2) \\
&= \frac{1}{(2\pi i)^2}\int\int_{B_{12}^2} \frac{f(\overset{(1)}{\phi_{12}}, \overset{(2)}{\phi_{12}})(\overset{2}{r} e^{i(\lambda_1+\lambda_2)})d\lambda_1 d\lambda_2}{(re^{i\lambda_1} - t_2)(re^{i\lambda_2} + C_2t_2 - t_1)} \\
(4.6)\qquad &+ \frac{1}{(2\pi i)^2}\int\int_{B_{13}^2} \frac{f(\overset{(1)}{\phi_{13}}, \overset{(2)}{\phi_{13}})(\overset{2}{r} e^{i(\lambda_1+\lambda_3)})d\lambda_1 d\lambda_3}{(re^{i\lambda_1} - t_2)(re^{i\lambda_3} - C_3t_2 - t_1)} \\
&+ \frac{1}{(2\pi i)^2}\int\int_{B_{23}^2} \frac{f(\overset{(1)}{\phi_{23}}, \overset{(2)}{\phi_{23}})(\overset{2}{r} e^{i(\lambda_2+\lambda_3)})d\lambda_2 d\lambda_3}{(re^{i\lambda_2} + C_2t_2 - t_1)(re^{i\lambda_3} - C_3t_2 - t_1)}.
\end{aligned}
$$

As in §2 we have that

$$
(4.7)\qquad a_{mn} = \frac{\partial^{m+n} f(0, 0)}{m!n!\partial t_1^m \partial t_2^n} = \frac{\partial^{m+n}[M_{12}(t_1, t_2) + M_{13}(t_1, t_2) + M_{23}(t_1, t_2)]}{m!n!\partial t_1^m \partial t_2^n}\bigg|_{t_1, t_2 = 0},
$$

$$
\begin{aligned}
&\frac{\partial^{m+n} M_{12}(t_1, t_2)}{\partial t_1^m \partial t_2^n} \\
(4.8)\qquad &= \frac{m!}{(2\pi i)^2}\int\int_{B_{12}^2} \frac{\partial^n}{\partial t_2^n}\left[\frac{f(\overset{(1)}{\phi_{12}}, \overset{(2)}{\phi_{12}})(\overset{2}{r} e^{i(\lambda_1+\lambda_2)})}{(re^{i\lambda_1} - t_2)(re^{i\lambda_2} + c_2t_2 - t_1)^{m+1}}\right] d\lambda_1 d\lambda_2 \\
&= \frac{m!n!}{(2\pi i)^2}\int\int_{B_{12}^2} \frac{f(\overset{(1)}{\phi_{12}}, \overset{(2)}{\phi_{12}})(\overset{2}{r} e^{i(\lambda_1+\lambda_2)})d\lambda_1 d\lambda_2}{[-\Phi_1(t_2)]^{n+1}[-\Phi_2(t_1, t_2)]^{m+1}},
\end{aligned}
$$

so that

$$
\begin{aligned}
&\frac{\partial^{m+n} M_{12}(0, 0)}{\partial t_1^m \partial t_2^n} = \frac{m!n!}{(2\pi i)^2}\int\int_{B_{12}^2} \frac{f(\overset{(1)}{\phi_{12}}, \overset{(2)}{\phi_{12}})(\overset{2}{r} e^{i(\lambda_1+\lambda_2)})}{r^{m+n+2}e^{i(m\lambda_2+n\lambda_1+\lambda_2+\lambda_1)}} \\
(4.9)\qquad &\cdot \sum_{\nu=0}^{n} \frac{(m+\nu)!}{m!\nu!}(c_2 e^{i(\lambda_1-\lambda_2)})^\nu d\lambda_1 d\lambda_2.
\end{aligned}
$$

This yields, by a process analogous to that used in §2,

$$
(4.10)\qquad \frac{1}{m!n!}\left|\frac{\partial^{m+n} M_{12}(0, 0)}{\partial t_1^m \partial t_2^n}\right| \leqq \frac{B_{12}(\overset{3}{g_{12}})M(r)G_{12}(m, n)}{r^{m+n}},
$$

where $M(r)$ is the maximum-modulus of f on $G^2(r)$, $B(g_{12}^3)$ is a constant depending on the hypersurface $g_{12}^3 = S_{r-r_0}^{r_1}G_{12}^2(r)$ and $G_{12}(m, n)$ is a function of m and n, also depending on g_{12}^3 and is defined in a way similar to $G(m, n)$ of §2.

In the same way we obtain

$$
(4.11)\qquad \frac{1}{m!n!}\left|\frac{\partial^{m+n} M_{13}(0, 0)}{\partial t_1^m \partial t_2^n}\right| \leqq \frac{B_{13}(\overset{2}{G_{13}})M(r)G_{13}(m, n)}{r^{m+n}}.
$$

From (4.3), we have

$$\Phi_2(t_1, t_2) = (t_1 - re^{i\lambda_2} - C_2t_2), \qquad \Phi_3(t_1, t_2) = (t_1 - re^{i\lambda_3} + C_3t_2). \tag{4.12}$$

Hence

$$\frac{\partial^m}{\partial t_1^m}\left[\frac{1}{\Phi_2\Phi_3}\right] = (-1)^m m! \sum_{\nu=0}^{m} \frac{1}{\Phi_2^{\nu+1}\Phi_3^{m-\nu+1}}, \tag{4.13}$$

and

$$\frac{\partial^m}{\partial t_1^m}\left[\frac{1}{\Phi_2\Phi_3}\right] = (-1)^{m+n} m!n! \sum_{\nu=0}^{m}\sum_{\mu=0}^{n} \frac{(m+n-\nu-\mu)!C_2^{\mu}C_3^{n-\mu}}{(m-\nu)!(n-\mu)!\Phi_2^{\nu+\mu+1}\Phi_3^{m+n-\nu-\mu+1}} \tag{4.14}$$

$$\begin{aligned} &= \frac{(-1)^{m+n}m!n!}{\Phi_3^{m+n+2}} \sum_{\mu=0}^{n} \frac{1}{(n-\mu)!}\left(\frac{C_2}{C_3}\,\frac{\Phi_3}{\Phi_2}\right)^{\mu} \\ &\quad \cdot \sum_{\nu=0}^{m} \frac{(m+n-\nu-\mu)!}{(m-\nu)!}\left(\frac{\Phi_3}{\Phi_2}\right)^{\nu}. \end{aligned} \tag{4.15}$$

Then

$$\begin{aligned} \frac{1}{m!n!}\left|\frac{\partial^{m+n}M_{23}(0, 0)}{\partial t_1^m \partial t_2^n}\right| &\leqq \frac{1}{4\pi^2}\iint_{B_{23}^2} \frac{\left|f(\overset{(1)}{\phi}_{23}, \overset{(2)}{\phi}_{23})\right|}{r^{m+n}} \sum_{\mu=0}^{n}\frac{1}{(n-\mu)!}\left(\frac{C_2}{C_3}\right)^{\mu} \\ &\quad \cdot \sum_{\nu=0}^{m}\frac{(m+n-\nu-\mu)!}{(m-\nu)!}\, d\lambda_2 d\lambda_3 \\ &\leqq \frac{B_{23}(\overset{2}{G}_{23})M(r)}{r^{m+n}} \sum_{\mu=0}^{n}\frac{1}{(n-\mu)!}\left(\frac{C_2}{C_3}\right)^{\mu} \\ &\quad \cdot \sum_{\nu=0}^{m}\frac{(m+u-\nu-\mu)!}{(m-\nu)!}. \end{aligned} \tag{4.16}$$

The constant $B_{23}(g_{23}^3)$ is given by $(1/4\pi^2)\iint_{B_{23}^2} d\lambda_2 d\lambda_3$, where the precise limits of integration are obtained by a tedious process and can be omitted here since they are not necessary for our purpose; it may be noted, however, that $0<B_{23}<1$. We shall denote by $G_{23}(m, n)$ the expression

$$\sum_{\mu=0}^{n}\frac{1}{(n-\mu)!}\left(\frac{C_2}{C_3}\right)^{\mu}\sum_{\nu=0}^{m}\frac{(m+n-\nu-\mu)!}{(m-\nu)!}. \tag{4.17}$$

This gives

$$\begin{aligned} |a_{mn}| &\leqq \frac{1}{m!n!}\left(\left|\frac{\partial^{m+n}M_{12}(0, 0)}{\partial t_1^m \partial t_2^n}\right| + \left|\frac{\partial^{m+n}M_{13}(0, 0)}{\partial t_1^m \partial t_2^n}\right| + \left|\frac{\partial^{m+n}M_{23}(0, 0)}{\partial t_1^m \partial t_2^n}\right|\right) \\ &\leqq \frac{M(r)}{r^{m+n}}\,\frac{1}{2}\sum_{k,s=1}^{3}{}' B_{ks}(\overset{3}{g}_{ks})G_{ks}(m, n). \end{aligned} \tag{4.18}$$

Therefore

$$M(r) \geqq \frac{r^{m+n} \mid a_{mn} \mid}{\frac{1}{2}\sum'^{3}_{k,s=1} B(g_{ks})G_{ks}(m, n)}. \tag{4.19}$$

Those values of μ and ν which make the right-hand side of (4.19) a maximum can be obtained by a process similar to that employed in §2. Hence

$$M(r) \geqq \frac{r^{\nu+\mu} \mid a_{\mu\nu} \mid}{\frac{1}{2}\sum'^{3}_{k,s=1} B_{ks}(g^{3}_{ks})G_{ks}(\mu, \nu)}. \tag{4.20}$$

We can then state

THEOREM III. *Given a function*

$$f(z_1, z_2) = \sum_{m,n=0}^{\infty} a_{mn} z_1^m z_2^n$$

regular in the closed domain $\overline{\mathfrak{M}}^4(r)$; *then along*

$$g^3 = \mathop{\mathrm{S}}_{r=r_0}^{r_1} G^2(r) = \mathop{\mathrm{S}}_{r=r_0}^{r_1} (G^2_{12}(r) + G^2_{13}(r) + G^2_{23}(r)),$$

$$M(r) \geqq \frac{r^{\nu+\mu} \mid a_{\mu\nu} \mid}{\frac{1}{2}\sum'^{3}_{k,s=1} B_{ks}(g^{3}_{ks})G_{ks}(\mu, \nu)}.$$

5. **Properties for the function on certain classes of surfaces lying in the boundary and different from $G^2(r)$.** We next wish to consider the growth of the function $f(z_1, z_2)$ over a special class of surfaces $H^2(r)$ belonging to the boundary of $\mathfrak{M}^4$. Let

$$H^2(r) \equiv E[z_1 = \zeta(r, \lambda_1, \sigma),\ z_2 = re^{i\lambda_1},\ \lambda_1^{(1)} \leqq \lambda_1 \leqq \lambda_1^{(2)},\ \sigma_1 \leqq \sigma \leqq \sigma_2] \tag{5.1}$$

where for all σ satisfying $\sigma_1 \leqq \sigma < \sigma_2$, and for any fixed λ_1 in the range considered,

$$\begin{aligned} &\zeta(r, \lambda_1, \sigma) \in \mathfrak{Z}^2_1(r, \lambda_1), \\ &\mathfrak{Z}^2_1(r, \lambda_1) \equiv A^2_1(r, \lambda_1) \cdot A^2_2(r, \lambda_1); \\ &A^2_1(r, \lambda_1) \equiv E[\mid z_1 - C_2 z_2 \mid \leqq r,\ z_2 = re^{i\lambda_1}], \\ &A^2_2(r, \lambda_1) \equiv E[\mid z_1 + C_3 z_2 \mid \leqq r,\ z_2 = re^{i\lambda_1}]; \end{aligned} \tag{5.2}$$

and for $\sigma = \sigma_2$, with λ_1 again fixed,

$$\zeta(r, \lambda_1, \sigma_2) \in s^1_1(r, \lambda_1),$$

where s_1 is the boundary of $\mathfrak{Z}^2_1(r, \lambda_1)$. It will be assumed that the set of all points of $H^2(r)$ for which λ_1 has an arbitrary fixed value in the range consid-

ered is a continuous curve $h^1(r)$[4] with an initial point $z_1=\zeta(r, \lambda_1, \bar\sigma_1(\lambda_1))$, and a terminal point on $s_1^1(r, \lambda_1)$. The surface $H^2(r)$ lies completely in that part of $s_1^3(r)$ which belongs to the boundary of the $\mathfrak{M}^4(r)$ of the previous section. A portion of the boundary of $H^2(r)$ lies on $G^2(r)$ of (4.2).

Let the maximum-modulus of $f(z_1, z_2)$ on $H^2(r)$ be $\gamma(r)$. We now map (using for simplicity the same notation for the mapped region) $\mathfrak{Z}_1^2$ into the unit circle so that $z_1=0$ goes into itself and the direction of the real axis, $\Re(z_1)=0$, at the point $z_1=0$, remains unchanged. The curve $h^1(r)$ maps into a segment of a continuous curve, its initial point determined by $\sigma=\bar\sigma_1(\lambda_1)$ and its terminal point lies on the unit circle. Now let $\theta=|z_1|=|\zeta(r, \lambda_1, \bar\sigma_1(\lambda_1))|$ for $\lambda_1^{(1)}\leqq\lambda_1\leqq\lambda_1^{(2)}$. The quantities θ and $\alpha=\lambda_1^{(2)}-\lambda_1^{(1)}$ were introduced by Bergman and are the characteristic numbers of the surface[5].

One form of the Milloux theorem is[6]: Let J be a continuous finite arc lying in the unit circle $|z|\leqq 1$ joining a point z_0 within the circle to a point on the boundary. Let $W(z)$ be regular, single-valued, and $|W(z)|<1$ inside the unit circle, and let $|W(z)|\leqq\omega$ on J. Then

$$|W(0)| < \omega^{(2/\pi)\sin^{-1}(1-\theta')/(1+\theta')}, \tag{5.3}$$

where $\theta'=|z_0|$.

Using this theorem for the mapped region $\mathfrak{Z}_1^2$ with

$$W(z_1) = \frac{f(z_1, z_2^*)}{M(r)}, \tag{5.4}$$

we have

$$|W(z_1)| = \frac{|f(z_1, z_2^*)|}{M(r)} \leqq \frac{\gamma(r)}{M(r)} < 1, \tag{5.5}$$

and get, letting $\Theta=(2/\pi)\sin^{-1}(1-\theta)/(1+\theta)$,

$$|f(0, z_2)| < M^{1-\Theta}\gamma^{\Theta}, \tag{5.6}$$

where $M(r)$ is the maximum-modulus of $f(z_1, z_2)$ in $\overline{\mathfrak{M}}^4$, λ_1^* is an arbitrarily chosen value of λ_1 in the range considered, and $z_2=z_2^*=re^{i\lambda_1^*}$.

Now

$$\begin{aligned} a_{0n} = \frac{1}{2\pi i}\Bigg\{\int_0^{\lambda_1^{(1)}} \frac{f(0, re^{i\lambda_1})}{r^n e^{in\lambda_1}}\,d\lambda_1 + \int_{\lambda_1^{(1)}}^{\lambda_1^{(2)}} \frac{f(0, re^{i\lambda_1})}{r^n e^{in\lambda_1}}\,d\lambda_1 \\ + \int_{\lambda_1^{(2)}}^{2\pi} \frac{f(0, re^{i\lambda_1})}{r^n e^{in\lambda_1}}\,d\lambda_1\Bigg\}, \end{aligned} \tag{5.7}$$

---

(4) The restriction that $h^1(r)$ be continuous is not essential since theorems of the Milloux type hold for more general one-dimensional sets.

(5) Bergman [1, pp. 347–348, Corollary]; and [4, pp. 200–201].

(6) R. Nevanlinna [5].

$$(5.8)\qquad |a_{0n}| \leqq \frac{1}{2\pi}\left\{\int_0^{\lambda_1^{(1)}} \frac{|f(0, re^{i\lambda_1})|}{r^n}\,d\lambda_1 + \int_{\lambda_1^{(1)}}^{\lambda_1^{(2)}} \frac{|f(0, re^{i\lambda_1})|}{r^n}\,d\lambda_1 + \int_{\lambda_1^{(2)}}^{2\pi} \frac{|f(0, re^{i\lambda_1})|}{r^n}\,d\lambda_1\right\},$$

and

$$(5.9)\qquad |a_{0n}| \leqq \frac{1}{2\pi r^n}\left[(2\pi - \alpha)\overline{M} + \alpha M\left(\frac{\gamma}{M}\right)^{\Theta}\right],$$

where $\overline{M} = \max |f(0, z_2)| = \max |\sum_{\nu=0}^{\infty} a_{0\nu} z_2^{\nu}|$. Then

$$(5.10)\qquad \frac{2\pi}{\alpha}\left[\frac{|a_{0n}|\,r^n}{M} - \frac{\overline{M}}{M}\right] + \frac{\overline{M}}{M} \leqq \left(\frac{\gamma}{M}\right)^{\Theta}.$$

Let Δ be defined by the equation

$$(5.11)\qquad \Delta = 1 - \frac{|a_{0\mu}|\,r^{\mu}}{\overline{M}},$$

μ being that n which maximizes $|a_{0n}|r^n$, and μ depends on r. Then Δ is positive.

If $\alpha > 2\pi\Delta$, we have that

$$(5.12)\qquad \gamma(r) \geqq M(r)\left[\frac{\overline{M}}{M}\left(1 - \frac{2\pi}{\alpha}\Delta\right)\right]^{\Theta^{-1}(\theta)},$$

where the right-hand side is positive.

From these results we can state

Theorem IV. *Given the function*

$$f(z_1, z_2) = \sum_{m,n=0}^{\infty} a_{mn} z_1^m z_2^n,$$

regular in $\overline{\mathfrak{M}}^4(r)$. *Let* $\max |f(z_1, z_2)| \leqq \gamma(r)$ *on the surface* $H^2(r)$ *of* (5.1) *having the characteristic numbers* $\theta(r)$ *and* α, $\alpha = \lambda_1^{(2)} - \lambda_1^{(1)} > 2\pi\Delta$, *where* $\Delta \equiv 1 - |a_{0\mu}|\,r^{\mu}/\overline{M}$; *then*

$$(5.13)\qquad \gamma(r) \geqq M(r)\left[\frac{\overline{M}(r)}{M(r)}\left(1 - \frac{2\pi\Delta}{\alpha}\right)\right]^{\Theta^{-1}(\theta)},$$

where $M = \max |f(z_1, z_2)|$ *and* $\overline{M} = \max |f(0, z_2)|$.

Since

$$(5.14)\qquad \frac{|a_{0\mu}|\,\rho^{\mu+1}}{\rho - r} > \overline{M}(r), \qquad \rho > r,$$

a lower bound for $\gamma(r)$ can be obtained in terms of the coefficients of $f(z_1, z_2)$ by replacing in (5.12) $M(r)$ by the right-hand side of (4.20), $\overline{M}(r)$ by $|a_{0\mu}|r^{\mu}$, and the $\overline{M}(r)$ in Δ by

$$\frac{|a_{0\mu'}|\rho_1^{\mu'+1}}{\rho_1 - r} \tag{5.15}$$

where

$$|a_{0\mu'}|\rho_1^{\mu'} = \underset{1+\epsilon<\rho<\infty}{\text{l.u.b.}} [\max_n |a_{0n}|\rho^n], \tag{5.16}$$

for an arbitrary positive ϵ. ρ_1 is a function of r and of the coefficients $\{a_{0n}\}$, and can be determined by a process similar to the Newton polygon method.

References

1. St. Bergman, *Zwei Sätze aus dem Ideenkreis des Schwarzschen Lemmas über die Funktionen von zwei komplexen Veränderlichen*, Mathematische Annalen, vol. 109 (1934), pp. 324–348.

2. ———, *Über eine in gewissen Bereichen mit Maximumfläche gültige Integraldarstellung der Funktionen zweier komplexer Variabler*, Mathematische Zeitschrift, vol. 39 (1934), pp. 76–79, 605–608.

3. ———, *Über eine Integraldarstellung von Funktionen zweier komplexer Veränderlichen*, Recueil Mathématique (Mathematičeski Sbornik), vol. 1 (43) (1936), pp. 851–862.

4. ———, *Über eine Abschätzung von meromorphen Funktionen zweier komplexer Veränderlicher in Bereichen mit ausgezeichneter Randfläche*, Travaux de l'Institut Mathématique de Tbilissi, vol. 1 (1937), pp. 187–204.

5. R. Nevanlinna, *Eindeutige analytische Funktionen*, chap. 4, §5.

Massachusetts Institute of Technology,
Cambridge, Mass.

ON THE POSSIBLE RATE OF GROWTH OF AN ANALYTIC FUNCTION

BY

P. W. KETCHUM

Introduction(1). The present paper deals with a number of diverse topics, ranging from purely topological considerations, through a general theory of possible distributions of values of an analytic function, to more special theorems on simultaneous expansions of an infinity of analytic functions. No unifying principle is presented as an excuse for treating such a variety of subjects; but there is a slight sequence of argument running throughout. The parts of the paper were actually written in reverse order. The initial investigation (Part III) was started as an attempt to generalize, from n to infinity, a known theorem(2) on simultaneous expansions of n analytic functions. This generalization was found to depend on an affirmative answer to the following question on level curves of an analytic function: Given any sequence of points $a_1, a_2, \cdots$, in the complex plane, which has the point at infinity as its only limit point, does there exist an analytic function with a level curve C such that C contains a distinct branch about each given point which separates that point from all the other points? This was, in turn, made to depend on a certain problem relative to the possible rate of growth of an integral function.

It was shown long ago by Poincaré(3), Borel(4), and others that an integral function may be made to grow arbitrarily fast along the real axis or along other lines or curves extending to infinity. Our problem was to obtain an affirmative answer to the following related question: Does there exist a sequence of regions $S_1, S_2, \cdots$, with a_i interior to S_i, such that, no matter how fast the sequence of numbers $m_1, m_2, \cdots$ increases there will be an integral function $f(z)$ for which

$$|f(t)| \geqq m_i \quad \text{in} \quad S_i?$$

In Part II we have shown that this is actually the case; but our construction will ordinarily give a function $f(z)$ with may zeros. This makes our desired

Presented to the Society, December 29, 1936, under the title *On the possible behavior of an analytic function at a set of isolated points*; received by the editors August 25, 1939, and, in revised form, June 3, 1940.

(1) Most of this paper was written while the writer was a member of the Institute for Advanced Study.

The writer is greatly indebted to Professors Bochner and Bohnenblust for valuable suggestions.

(2) See §11 of this paper.

(3) American Journal of Mathematics, vol. 14 (1892), p. 214.

(4) *Leçons sur les Séries à Termes Positifs*, Paris, 1902, p. 27.

level curve C have not only the required branches, but many extraneous ones in addition. Whether or not there exists a nonvanishing function $f(z)$ with the stated properties, so that the corresponding level curve C has exactly the required branches and no more, is a question that we have been unable to settle.

On investigating what sort of regions $S_1, S_2, \cdots$ may be used to obtain the required properties, certain questions concerning the topology on a sphere arose; these are considered in Part I. In so doing, it was found expedient to make certain restatements of Hilbert's theorem on the approximation of Jordan curves by lemniscates.

Part I

1. **Preliminary definitions.** By a *ring* on the extended complex plane or the Riemann sphere, we mean an open or closed region bounded by two Jordan curves that have no point in common. The point at infinity may be interior to, exterior to, or on the boundary of the ring.

The only sets or sequences of rings considered in this paper are those in which the rings are *mutually non-intersecting*; that is, no two of the rings have closures with points in common. For this reason we shall usually not bother to state explicitly the non-intersecting character of the rings, but this condition is always to be understood.

Topologically, there is no distinction between a given ring on a sphere and any other ring on a sphere. Likewise every pair of rings is equivalent to every other pair; the complementary regions into which the rings divide the plane necessarily consist of two simply connected regions A and B and a third ring R which separates A from B. For three or more rings, on the other hand, we may distinguish relative positions according to the arrangement and connectivity of the complementary regions. If there are n given rings, the connectivity k of the complementary region of maximum connectivity may vary from 2 to n. We distinguish the two extreme cases by names: if k is 2 we call the rings *nested*, if $k=n$ the rings are *mutually exterior*. (This must not be confused with mutually exterior in the point-set-theoretic sense, which means non-intersecting in our terminology.)

The multiply connected region which is complementary to a set of n mutually exterior rings will be called the *R-exterior* of the rings; the other complementary regions, all simply connected and n in number, the *R-interiors*.

The same definitions of nested, mutually exterior, R-exterior, and R-interior will evidently apply as well to finite sets of non-intersecting Jordan curves on the sphere.

2. **Generalizations of Hilbert's theorem.** According to Walsh and Russell's(5) generalization of a theorem due to Hilbert, a finite number of mu-

(5) These Transactions, vol. 36 (1934), pp. 13–28.

tually exterior Jordan curves whose R-exterior contains the point at infinity can be uniformly approximated by means of lemniscates.

As thus stated, Hilbert's theorem imposes a special rôle on the point at infinity. To avoid this, we generalize the notion of lemniscate as follows: A *lemniscate with poles at* $b_1, \cdots, b_n$ *and zeros at* $a_1, \cdots, a_m$ is a level curve of a rational function whose poles are at $b_1, \cdots, b_n$ and whose zeros are at $a_1, \cdots, a_m$.

By imposing the transformation $z=1/(z'-b)$ in Hilbert's theorem, we get the following more symmetrical statement: Any set of mutually exterior Jordan curves can be uniformly approximated by a lemniscate with pole at b and zeros at $a_1, \cdots, a_m$, where b is a preassigned point in the R-exterior and the a's are somewhere in the R-interiors of the given curves.

This statement suggests a dual theorem in which the rôles of pole and zeros are interchanged. To get such a dual theorem we may first consider the reciprocal of the rational function yielding the lemniscate in the last statement. It follows that the given Jordan curves can be approximated by a lemniscate with zero at b and poles at $a_1, \cdots, a_m$. Now, applying Runge's method of moving of poles, we may approximate our rational function uniformly in the R-exterior of the curves by another rational function with poles at given points $b_1, \cdots, b_n$, one in each of the R-interiors of the curves. The degree of approximation being arbitrary, one may also choose this new function so that its zeros lie as close as one pleases to b. Repeating this process, we obtain the following four theorems, of which Theorem I is a sharpening of Hilbert's theorem.

THEOREM I [THEOREM I′]. *Let $R_1, \cdots, R_n$ be given mutually exterior rings on the extended complex plane, and let b be a given point in the R-exterior of the rings and $a_1, \cdots, a_n$ given points such that a_i is in the R-interior of R_i. Then there is a lemniscate, whose only pole [zero] is at b, which consists of n Jordan curves, $C_1, \cdots, C_n$, with C_i interior to R_i (in the point-set-theoretic sense) and with C_i separating the R-interior of R_i from its R-exterior. Furthermore, the zeros [poles] of the lemniscate may be taken at a set of points $c_{i,j}$, $i=1, \cdots, n$, $j=1, \cdots, l_i$, such that the distance from any $c_{i,j}$ to a_i is less than ϵ, where ϵ is an arbitrarily small positive number.*

THEOREM II [THEOREM II′]. *Under the same hypotheses as in Theorem* I, *there is a lemniscate with poles [zeros] at $a_1, \cdots, a_n$ which consists of n Jordan curves $C_1, \cdots, C_n$ with the same properties as in Theorem* I. *Furthermore, the zeros [poles] of the lemniscate can be taken at points $c_1, \cdots, c_l$, where each c_i is at distance less than ϵ from b.*

3. Further definitions. A point b will be called a *sequential limit point* of a sequence of rings $R_1, R_2, \cdots$ if there is a sequence of points $\beta_1, \beta_2, \cdots$, with β_i on R_i, which has b as a limit point.

More generally, if $\{S_\nu\}$ is an infinite collection of mutually disjunct sets, either denumerable or nondenumerable in number, then a point b will be called a *sequential limit point of the collection* if there is a sequence of points $\beta_1, \beta_2, \cdots$ such that (i) every β_i is in some S_ν, (ii) no two of the β's are in the same S_ν, (iii) the sequence of β's has b as a limit point.

It can be shown that the rings $R_1, R_2, \cdots, R_n$ are nested if and only if they can be ordered in such a way that R_i separates R_j from R_k whenever i is between j and k.

An *infinite sequence of rings* will be called *nested* if it can be arranged in an order $R_1, R_2, \cdots$ such that R_i separates R_j from R_k whenever i is between j and k. An equivalent definition is that, when properly ordered, R_j separates R_k from the set of sequential limit points whenever j is greater than k.

We shall denote the boundary of a set A by $F(A)$ and its complement by $C(A)$.

An *interior sequential limit point* of a sequence of rings is a sequential limit point of a nested subsequence of the rings.

Every closed set B, not the null set or the whole plane, divides the plane into a finite number or denumerable infinity of domains $\sigma_1, \sigma_2, \cdots$ namely, the components of $C(B)$. A closed set B will be said to be *encased* in a sequence of rings in $C(B)$ if, for each i, those rings which are in σ_i form a subsequence with the following two properties:

(a) The set of sequential limit points, the interior sequential limit points, and $F(\sigma_i)$ are three identical sets.

(b) If $F(\sigma_i)$ has more than one component, then for every ring there are points of $F(\sigma_i)$ in both of the two regions complimentary to the ring. If $F(\sigma_i)$ has only one component, then the rings are nested.

4. A theorem of plane topology.

THEOREM III. *Let B be any closed set, not the null set or the whole plane, and let S be a set in $C(B)$ whose components $\{S_\nu\}$ are closed and have sequential limit points only in B. Then there exists a sequence of rings in $C(B+S)$ which encases B. The boundary curves of these rings may be taken to be lemniscates.*

Proof. We assume a metric for the entire extended plane, as, for instance, distances on the Riemann sphere.

For the proof it will be sufficient to show that if σ_i is any component of $C(B)$, then there exists in $\sigma_i - S\sigma_i$ a sequence of rings which satisfies the above condition (b) and whose sequential limit points and interior sequential limit points are both identical with $F(\sigma_i)$.

If z is a point in a non-null set A with a non-null boundary, we shall call its distance from $F(A)$ the *depth* of z in A. In any non-null set A there will be one or more points which have maximum depth $\rho_0 \geqq 0$ in A. In particular, if A is open, then there will be a point having maximum depth $\rho_0 > 0$. The number ρ_0 will be called the *maximum depth* of the set A. Since our metric is

bounded, every set has finite maximum depth. If ρ_0 is the maximum depth of A, then for every number ρ, $0\leqq\rho\leqq\rho_0$ there are points in A with depth ρ. If B is a closed set in A, we shall call the greatest depth of any point of B in A the *maximum depth of B in A* and the smallest depth of any point of B in A the *minimum depth of B in A*.

The components of S which have points in common with σ_i are entirely in σ_i. Let $K(\rho)$ be the sum of the components of S whose minimum depth in σ_i is larger than ρ, where ρ is any positive number less than the maximum depth ρ_0 of σ_i. There will be at most a finite number of components of S in $K(\rho)$: otherwise there would be a sequential limit point of the components at a positive depth in σ_i, which is contrary to hypothesis. Let $k(\rho)$ be the sum of the components of S whose minimum depth in σ_i is not greater than ρ. Let $A(\rho)$ be the set of all points whose depth in σ_i is greater than ρ. The boundary $F(A(\rho))$ is non-vacuous. Let $a(\rho)$ be the set of all points whose depth in σ_i is not greater than ρ.

We will now show that if $0<\rho_1<\rho_0$ then there is a number ρ_2, $0<\rho_2<\rho_1$, such that $a(\rho_2)+k(\rho_2)$ is at a positive distance, say δ, from $A(\rho_1)+K(\rho_2)$. Obviously $a(\rho_2)$ is at a positive distance from $A(\rho_1)$ and from $K(\rho_2)$, and $k(\rho_2)$ from $K(\rho_2)$. Suppose no number ρ_2 exists such that $k(\rho_2)$ is at a positive distance from $A(\rho_1)$. No fixed component of S can be in $k(\rho_2)$ for every ρ_2; hence there must exist a sequence of components $S_1, S_2, \cdots$ such that the distance from S_j to $A(\rho_1)$ approaches zero with j. But this would imply a sequential limit point of the S_j at positive depth in σ_i, which is contrary to hypothesis.

Cover the closure of $a(\rho_2)+k(\rho_2)$ by a finite number of circular closed regions with centers on this set and radius $\eta<\delta$. The number η can be chosen so that no two of the circular boundaries of these regions will be tangent to each other, and no three intersect at a single point. Let $L(\eta)$ be the sum of these circular regions.

If we start at any point on the boundary of $L(\eta)$ and proceed along the boundary, we will arrive back at the starting point after traversing a finite number of arcs of circles and an equal number of points where two circles intersect. The boundary of $L(\eta)$ consists of a finite number of Jordan curves. Furthermore, it is possible to increase the radius of each circle in $L(\eta)$ slightly without changing the connections between the circular arcs forming the boundary of $L(\eta)$. Otherwise, the connections would change in a discontinuous manner; but this can happen only when two arcs are tangent or when more than two arcs intersect at a point. Hence there is an η' between η and δ such that $[L(\eta')-L(\eta)]\sigma_i$ consists of a finite number of rings $R_1, R_2, \cdots, R_k$. By their construction, these rings have a minimum depth in σ_i greater than ρ_2, a maximum depth less than ρ_1, and are at a positive distance from S.

A ring R will be said to *bound* a set A, $AR\neq 0$, if one of the Jordan curves forming the boundary of R is contained in $F(A)$. A collection of rings $R_1, R_2, \cdots, R_p$ will be said to *completely bound* A if each of the rings bounds

A and if $F(A)$ is contained in the closure of the rings. A set of rings which completely bounds a connected set will be mutually exterior and will separate A from $C(A+\sum R_i)$.

The rings $R_1, R_2, \cdots, R_k$ will completely bound $\sigma_i - L(\eta')\sigma_i$. Let M be a component of $\sigma_i - L(\eta')\sigma_i$ which contains a point x_0 of maximum depth ρ_0 in σ_i. There will be a subset of the rings, $R_1, R_2, \cdots, R_p$, which completely bound M. These rings will be mutually exterior and will separate M from $F(\sigma_i)$.

If x is any point in σ_i, it can be joined to x_0 by a Jordan arc in σ_i. This arc will have a positive distance λ from $F(\sigma_i)$, that is, it will have positive minimum depth λ in σ_i. Hence there exists a ρ_1 so small that the corresponding set M will contain x, and the corresponding rings $R_1, \cdots, R_p$ will separate x from $F(\sigma_i)$. If $F(\sigma_i)$ is connected, then M will be simply connected for every ρ_1. On the other hand, if $F(\sigma_i)$ is not connected, then M will be multiply connected for every sufficiently small ρ_1.

Let ϵ_1 be less than ρ_0 and also so small that if $F(\sigma_i)$ is not connected then the set M corresponding to any $\rho_1 \leqq \epsilon_1$ will be multiply connected. Starting with ϵ_1 we form any sequence $\epsilon_1, \epsilon_2, \cdots$ of positive numbers with zero as a limit and such that if ϵ_j is taken to be ρ_1 in the above, then ρ_2 may be taken equal to ϵ_{j+1}. That is, $a(\epsilon_{j+1})+k(\epsilon_{j+1})$ is at a positive distance from $A(\epsilon_j)+K(\epsilon_{j+1})$. Corresponding to each pair of values $\epsilon_j, \epsilon_{j+1}$ there will be a set of rings $R_1^{(j)}, \cdots, R_p^{(j)}$ constructed as above on taking $\rho_1=\epsilon_j$ and $\rho_2=\epsilon_{j+1}$.

It will now be shown that the totality of these rings is a sequence which encases $F(\sigma_i)$.

(i) That condition (b) will be satisfied is a consequence of the fact that the rings $R_1^{(j)}, \cdots, R_p^{(j)}$ completely bound a set M_j which is simply connected if $F(\sigma_i)$ has only one component and multiply connected otherwise.

(ii) Every sequential limit point of the rings is on $F(\sigma_i)$ because $R_k^{(j)}$ has maximum depth less than ϵ_j and $\epsilon_j \to 0$.

(iii) Every point of $F(\sigma_i)$ is a sequential limit point of the rings. If not, then there exists a point x in $F(\sigma_i)$ and a circular neighborhood N_x such that N_x is free of points of the rings. But N_x contains at least one point y in σ_i. For j sufficiently great, $R_k^{(j)}$ separates y from x. This can happen only if $R_k^{(j)}$ has points in N_x, which is a contradiction.

(iv) It remains to show that every point of $F(\sigma_i)$ is an interior sequential limit point of the rings; that is, for every x in $F(\sigma_i)$ there is a nested subsequence of the rings which has x as a sequential limit point. Since the rings $R_1^{(j)}, \cdots, R_p^{(j)}$ separate x_0 from $F(\sigma_i)$, there will be a particular ring, say $R_{\mu_j}^{(j)}$, which separates x_0 from x. The sequence of rings $\{R_{\mu_j}^{(j)}\}$ will have x as a sequential limit point and will be nested.

That the boundary curves of the rings can be taken to be lemniscates follows immediately from Theorem II and the fact that the rings $R_1^{(j)}, \cdots, R_p^{(j)}$ are mutually exterior. In fact, the poles of the lemniscate may be taken on $F(\sigma_i)$, since if $p>1$ each ring contains a point of $F(\sigma_i)$ in its R-interior.

PART II

5. Construction of functions with very rapid growth.

THEOREM IV. *Let $G(z) \geqq 0$ be a given function defined for every point z in the extended complex plane. Let B be the set consisting of all points in every neighborhood of which $G(z)$ is unbounded. We suppose that B is neither the null set nor the whole extended plane. Let $R_1, R_2, \cdots$ be any sequence of mutually nonintersecting rings which encases B. Then there exists a function $f(z)$ which is analytic except possibly at points of B and which satisfies the inequality*

$$|f(z)| \geqq G(z) \tag{1}$$

for every z not in one of the rings.

COROLLARY. *Let $G(z)$ and B be given as in Theorem* IV. *Let S be any set in $C(B)$ whose components are closed and whose sequential limit points are all on B. Then there is a function $f(z)$ which is analytic except on B and which satisfies the inequality* (1) *at all points of S.*

By Theorem III, this corollary is an immediate consequence of Theorem IV.

In this section and throughout the paper, all functions are assumed to be single valued.

In case $G(z)$ is bounded in the whole plane (B is the null set), the problem is trivial; since one can then satisfy (1) everywhere by taking $f(z)$ to be a sufficiently large constant.

Proof of Theorem IV. We confine our attention to one of the domains σ_k into which B divides the plane, and to those rings which lie in σ_k. Let P be any point in σ_k and not in any of the rings. Then we distinguish the two parts of the plane exterior to any ring R_j by the *inside* and *outside* of R_j according to whether it does or does not contain P, respectively. Having chosen P, every ring has an inside and outside, but different choices of P may interchange the inside and outside of a given ring. We suppose that P is now chosen once for all, subject only to the condition that if the complement S of the rings in σ_k has a simply connected component S_0 (which can happen only if $F(\sigma_k)$ is connected), then P is in S_0.

If a nested subsequence of the rings in σ_k has a point x on $F(\sigma_k)$ as a sequential limit point, then any point y in σ_k will be separated from x by all but a finite number of this subsequence of rings. For, let y be a point in σ_k which is not separated from x by an infinity of rings. Let z be a point in σ_k which is separated from x by the first of the rings. Join z to y by a Jordan arc in σ_k. Every ring separates z from x. There must be an infinite number of rings which separate y from z. These rings will all intersect the Jordan arc yz; therefore, there will be a sequential limit point of the rings on the arc, and hence in σ_k; which is a contradiction.

A and if $F(A)$ is contained in the closure of the rings. A set of rings which completely bounds a connected set will be mutually exterior and will separate A from $C(A+\sum R_i)$.

The rings $R_1, R_2, \cdots, R_k$ will completely bound $\sigma_i - L(\eta')\sigma_i$. Let M be a component of $\sigma_i - L(\eta')\sigma_i$ which contains a point x_0 of maximum depth ρ_0 in σ_i. There will be a subset of the rings, $R_1, R_2, \cdots, R_p$, which completely bound M. These rings will be mutually exterior and will separate M from $F(\sigma_i)$.

If x is any point in σ_i, it can be joined to x_0 by a Jordan arc in σ_i. This arc will have a positive distance λ from $F(\sigma_i)$, that is, it will have positive minimum depth λ in σ_i. Hence there exists a ρ_1 so small that the corresponding set M will contain x, and the corresponding rings $R_1, \cdots, R_p$ will separate x from $F(\sigma_i)$. If $F(\sigma_i)$ is connected, then M will be simply connected for every ρ_1. On the other hand, if $F(\sigma_i)$ is not connected, then M will be multiply connected for every sufficiently small ρ_1.

Let ϵ_1 be less than ρ_0 and also so small that if $F(\sigma_i)$ is not connected then the set M corresponding to any $\rho_1 \leqq \epsilon_1$ will be multiply connected. Starting with ϵ_1 we form any sequence $\epsilon_1, \epsilon_2, \cdots$ of positive numbers with zero as a limit and such that if ϵ_j is taken to be ρ_1 in the above, then ρ_2 may be taken equal to ϵ_{j+1}. That is, $a(\epsilon_{j+1})+k(\epsilon_{j+1})$ is at a positive distance from $A(\epsilon_j)+K(\epsilon_{j+1})$. Corresponding to each pair of values $\epsilon_j, \epsilon_{j+1}$ there will be a set of rings $R_1^{(j)}, \cdots, R_p^{(j)}$ constructed as above on taking $\rho_1=\epsilon_j$ and $\rho_2=\epsilon_{j+1}$.

It will now be shown that the totality of these rings is a sequence which encases $F(\sigma_i)$.

(i) That condition (b) will be satisfied is a consequence of the fact that the rings $R_1^{(j)}, \cdots, R_p^{(j)}$ completely bound a set M_j which is simply connected if $F(\sigma_i)$ has only one component and multiply connected otherwise.

(ii) Every sequential limit point of the rings is on $F(\sigma_i)$ because $R_k^{(j)}$ has maximum depth less than ϵ_j and $\epsilon_j \to 0$.

(iii) Every point of $F(\sigma_i)$ is a sequential limit point of the rings. If not, then there exists a point x in $F(\sigma_i)$ and a circular neighborhood N_x such that N_x is free of points of the rings. But N_x contains at least one point y in σ_i. For j sufficiently great, $R_k^{(j)}$ separates y from x. This can happen only if $R_k^{(j)}$ has points in N_x, which is a contradiction.

(iv) It remains to show that every point of $F(\sigma_i)$ is an interior sequential limit point of the rings; that is, for every x in $F(\sigma_i)$ there is a nested subsequence of the rings which has x as a sequential limit point. Since the rings $R_1^{(j)}, \cdots, R_p^{(j)}$ separate x_0 from $F(\sigma_i)$, there will be a particular ring, say $R_{\mu_j}^{(j)}$, which separates x_0 from x. The sequence of rings $\{R_{\mu_j}^{(j)}\}$ will have x as a sequential limit point and will be nested.

That the boundary curves of the rings can be taken to be lemniscates follows immediately from Theorem II and the fact that the rings $R_1^{(j)}, \cdots, R_p^{(j)}$ are mutually exterior. In fact, the poles of the lemniscate may be taken on $F(\sigma_i)$, since if $p>1$ each ring contains a point of $F(\sigma_i)$ in its R-interior.

PART II

5. Construction of functions with very rapid growth.

THEOREM IV. *Let $G(z) \geqq 0$ be a given function defined for every point z in the extended complex plane. Let B be the set consisting of all points in every neighborhood of which $G(z)$ is unbounded. We suppose that B is neither the null set nor the whole extended plane. Let $R_1, R_2, \cdots$ be any sequence of mutually nonintersecting rings which encases B. Then there exists a function $f(z)$ which is analytic except possibly at points of B and which satisfies the inequality*

$$(1) \qquad |f(z)| \geqq G(z)$$

for every z not in one of the rings.

COROLLARY. *Let $G(z)$ and B be given as in Theorem* IV. *Let S be any set in $C(B)$ whose components are closed and whose sequential limit points are all on B. Then there is a function $f(z)$ which is analytic except on B and which satisfies the inequality* (1) *at all points of S.*

By Theorem III, this corollary is an immediate consequence of Theorem IV.

In this section and throughout the paper, all functions are assumed to be single valued.

In case $G(z)$ is bounded in the whole plane (B is the null set), the problem is trivial; since one can then satisfy (1) everywhere by taking $f(z)$ to be a sufficiently large constant.

Proof of Theorem IV. We confine our attention to one of the domains σ_k into which B divides the plane, and to those rings which lie in σ_k. Let P be any point in σ_k and not in any of the rings. Then we distinguish the two parts of the plane exterior to any ring R_j by the *inside* and *outside* of R_j according to whether it does or does not contain P, respectively. Having chosen P, every ring has an inside and outside, but different choices of P may interchange the inside and outside of a given ring. We suppose that P is now chosen once for all, subject only to the condition that if the complement S of the rings in σ_k has a simply connected component S_0 (which can happen only if $F(\sigma_k)$ is connected), then P is in S_0.

If a nested subsequence of the rings in σ_k has a point x on $F(\sigma_k)$ as a sequential limit point, then any point y in σ_k will be separated from x by all but a finite number of this subsequence of rings. For, let y be a point in σ_k which is not separated from x by an infinity of rings. Let z be a point in σ_k which is separated from x by the first of the rings. Join z to y by a Jordan arc in σ_k. Every ring separates z from x. There must be an infinite number of rings which separate y from z. These rings will all intersect the Jordan arc yz; therefore, there will be a sequential limit point of the rings on the arc, and hence in σ_k; which is a contradiction.

Let S be the set consisting of all points in σ_k which are not in any ring. Each component S_j of S is completely bounded by a finite number of the rings. For, if an infinite number of rings bound S_j, then there will be a sequential limit point x of the rings in the closure of S_j. This point x will also be on $F(\sigma_k)$ and hence will be a sequential limit point of a nested subsequence of the rings. Any point y of S_j is in σ_k. We have just seen that y is separated from x by all but a finite number of rings. But this is a contradiction since S_j+x is a connected set, in the complement of the rings, which contains x and y.

Denote the component of S which contains P by S_0. There will be only a finite number of rings which separate any component from P, and only a finite number of components, $S_{i1}, \cdots, S_{il_i}$, each of which is separated from P by exactly i rings. We now rename the rings by using a double subscript R_{ij}, $j=1, 2, \cdots, l_i$, so that S_{ij} is outside of and bounded by R_{ij}. Let I_{ij} be the inside of R_{ij}. Put

$$I^{(i)} = I_{i1}\cdot I_{i2} \cdots I_{il_i}, \qquad S^{(i)} = S_{i1} + S_{i2} + \cdots + S_{il_i}.$$

The rings $R_{i1}, \cdots, R_{il_i}$ are mutually exterior, since they completely bound the connected set $I^{(i)}$; and there will be points of $F(\sigma_k)$ outside each of these rings.

Let M_0 be the upper bound of $G(z)$ in S_0 and M_i the upper bound in $S^{(i)}$.

Let $\epsilon_1, \epsilon_2, \cdots$ be any set of positive numbers such that $\sum \epsilon_i$ converges to a given sum ϵ.

Let N_1 be the positive constant $M_0+\epsilon$. By Theorem II there is a rational function $F_1(z)$, analytic except at points of B, such that the lemniscate $|F_1(z)|=1$ consists of l_1 contours $C_{11}, \cdots, C_{1l_1}$, where C_{1j} is interior to R_{1j} and separates S_{1j} from P, $j=1, 2, \cdots, l_i$. Then there will exist a positive integer p_1 so large that

$$\begin{aligned} |F_1(z)|^{p_1} &< \epsilon_1 && \text{in} \quad S_0 = I^{(1)}, \\ &> M_1 + N_1 + \epsilon && \text{in} \quad S^{(1)}. \end{aligned}$$

Let N_2 be the upper bound of $|N_1+[F_1(z)]^{p_1}|$ in $S^{(2)}$.

Similarly, there is a rational function $F_2(z)$, whose only poles are on B, such that the lemniscate $|F_2(z)|=1$ consists of l_2 contours $C_{21}, \cdots, C_{2l_2}$, where C_{2j} is interior to R_{2j} and separates S_{2j} from P. Then there will be a positive integer p_2 so large that

$$\begin{aligned} |F_2(z)|^{p_2} &< \epsilon_2 && \text{in} \quad I^{(2)}, \\ &> M_2 + N_2 + \epsilon && \text{in} \quad S^{(2)}. \end{aligned}$$

Continuing this process, we suppose that $F_3(z), \cdots, F_{i-1}(z)$ have been constructed. Let

$$N_{i+1} \geqq | N_1 + [F_1(z)]^{p_1} + [F_2(z)]^{p_2} + \cdots + [F_{i-1}(z)]^{p_{i-1}} |$$

in $S^{(i)}$. There is a rational function $F_i(z)$, whose only poles are on B, such that the lemniscate $|F_i(z)|=1$ will consist of l_i contours $C_{i1}, C_{i2}, \cdots, C_{il_i}$, where C_{ij} is interior to R_{ij} and separates S_{ij} from P. Then there will be a positive integer p_i so large that

$$\begin{aligned} |F_i(z)|^{p_i} &< \epsilon_i \quad \text{in} \quad I^{(i)} \\ &> M_i + N_i + \epsilon \quad \text{in} \quad S^{(i)}. \end{aligned}$$

Now consider the function

$$f(z) = N_1 + \sum_{i=1}^{\infty} [F_i(z)]^{p_i}. \tag{2}$$

Let Σ be any closed region exterior to B, and let z be a point of Σ. From the manner in which the terms of (2) were constructed, it is clear that one can find an integer q so large that

$$|F_i(z)|^{p_i} < \epsilon_i \qquad \text{for } z \text{ in } \Sigma \text{ and } i > q.$$

Since $\sum \epsilon_i$ converges, this shows that the series (2) converges uniformly in Σ. Hence $f(z)$ is analytic everywhere except for the points of B.

Moreover, for z in one of the regions $S_{q1}, S_{q2}, \cdots, S_{ql_q}$, one can write

$$\begin{aligned} |f(z)| &\geqq |F_q(z)|^{p_q} - \left| N_1 + \sum_{i=1}^{q-1} [F_i(z)]^{p_i} \right| - \sum_{i=q+1}^{\infty} |F_i(z)|^{p_i} \\ &\geqq (M_q + N_q + \epsilon) - N_q - \sum_{i=q+1}^{\infty} \epsilon_i \\ &\geqq M_q \geqq G(z), \end{aligned}$$

and the theorem follows.

6. **Best possible character of Theorem** IV. In Theorem IV the rings $R_1, R_2, \cdots$ present a sort of barrier between the set B, where $G(z)$ is unbounded, and the set S, where (1) holds. It is true that points of B may be limit points of S, but no point of B can be a limit point of points in any single region in S. Thus, collectively the regions of S are close to B, but individually they are not.

The question arises as to whether some such barrier is necessary. We answer this question in the affirmative in the following theorem, which states that no theorem like Theorem IV can be true in case a region in S has a limit point on B.

THEOREM V. *Let B be any closed set, not the null set or the whole plane, and let S be an open region which has at least one point of B on its boundary, but has no points in common with B. Then there exists a function $G(z)$ such that B consists of precisely those points in every neighborhood of which $G(z)$ is unbounded,*

and such that no function $f(z)$ exists which is meromorphic in S and satisfies (1) *in S.*

Proof. Let $C_1, C_2, \cdots$ be a nested sequence of analytic Jordan curves in S, having a point b of B as an interior sequential limit point, and such that the open regions $\Sigma_1, \Sigma_2, \cdots$ in S bounded by these contours all have an interior point in common, say s_1. Then there exists a sequence of mapping functions which will map $\Sigma_1, \Sigma_2, \cdots$ conformally onto the interior of the unit circle in such a way that the origin is the image of s_1. The sequence of one-to-one transformations thus defined will transform any function $f(z)$, meromorphic in S, into a sequence of functions $\{f_i(z)\}$ which are meromorphic in and on the unit circle.

Let $\theta_1, \theta_2, \cdots$ be a sequence of arcs on the unit circle such that the corresponding sequence of arcs $\theta_1', \theta_2', \cdots$ on the curves $C_1, C_2, \cdots$ will have b as its only sequential limit point. For brevity, we also denote by $\theta_1, \theta_2, \cdots$ the lengths of the arcs $\theta_1, \theta_2, \cdots$. Let $M_1, M_2, \cdots$ be any sequence of positive constants such that $\theta_i \log M_i \to \infty$ as $i \to \infty$.

Let $G(z)$ be defined as equal to unity in S except on the arcs $\theta_1', \theta_2', \cdots$; where we take $G(z) = M_1, M_2, \cdots$, respectively. Outside S, $G(z)$ is given any values such that $G(z) \to \infty$ if and only if z tends to a point of B. The points B will be precisely those in every neighborhood of which this function $G(z)$ is unbounded.

We now show that no function $f(z)$ can exist which is meromorphic in S and satisfies (1) in S. Suppose, on the contrary, that such a function did exist. Then the transformed functions $f_i(z)$ would be meromorphic in and on the unit circle and $|f_i(z)| \geqq G_i(z)$ for $|z| \leqq 1$, where the $G_i(z)$ are the transforms of $G(z)$. Hence, by Jensen's theorem, since $|f_i(z)| \geqq 1$ for $|z| \leqq 1$,

$$\log |f_i(0)| = \frac{1}{2\pi}\int_0^{2\pi} \log |f_i(e^{i\theta})| \, d\theta + \sum_p \log \frac{1}{|b_{ip}|},$$

where the b_{ip} are the poles of $f_i(z)$ in the unit circle. The last term is not negative, and the integral is larger than $\theta_i \log M_i$; so that

$$\log |f_i(0)| = \log |f(s_1)| \geqq \frac{1}{2\pi} \theta_i \log M_i.$$

The right-hand member of this equation tends to infinity as i becomes infinite, while the left-hand member is fixed, independent of i. We thus have a contradiction, and the theorem is proved.

7. **Generalizations of Theorems** IV **and** V. The last two theorems may be generalized by replacing the inequality (1) by other relationships. We may, for instance, let $g(z)$ be a given analytic function and require that

$$(1') \qquad f(z) = g(z) + h(z), \qquad |h(z)| \leqq G(z).$$

This problem is one of approximation to analytic functions by means of analytic functions, the degree of approximation to be predetermined by the given function $G(z)$. We thus have

THEOREM IV′. *Let $G(z)$ be a given function which is defined and positive in the entire complex plane. Let B be the set of points in every neighborhood of which $G(z)$ has zero as a greatest lower bound. We assume that B is neither the null set nor the whole extended plane. Let $R_1, R_2, \cdots$ be any sequence of rings which encases B, and $g(z)$ any function which is analytic for all z not in B. Then there is a function $f(z) \not\equiv g(z)$ which is meromorphic except at points of B and which satisfies* (1′) *for every z not in one of the rings.*

THEOREM V′. *Let B be a given closed set, not the null set or the whole extended plane, and let S be an open region which has at least one point of B on its boundary, but has no points in common with B. Let $g(z)$ be a given function which is analytic for every z not in B. Then there exists a function $G(z)>0$ such that B consists of precisely those points in every neighborhood of which $G(z)$ has zero as a greatest lower bound, and such that no function $f(z)$ exists which is meromorphic in S and satisfies* (1′) *in S.*

These two theorems are obtained immediately when one applies Theorems IV and V, respectively, to the reciprocals of $h(z)$ and $G(z)$.

8. **An improvement of Theorem IV in a special case.** In the next theorem we present an example of a situation intermediate between Theorems IV and V. The theorem involves a certain kind of set S which resembles a cartwheel and which is described as follows: (a) The set S includes all points in an infinite sequence of concentric, circular rings with centers at the origin, which has infinity as its only sequential limit point. (b) Let $\alpha_1, \alpha_2, \cdots$ be a sequence of rays from the origin whose angles with the real axis are rational multiples of 2π, and let $r_1, r_2, \cdots$ be positive numbers with $r_i \to \infty$; then S includes all those points on α_i where $|z| \geqq r_i$, $i=1, 2, \cdots$.

THEOREM VI. *Let S be a given set of points of the sort just described. Let $G(z)$ be any given positive valued function which is bounded in every bounded region of the plane. Then there exists an integral function $f(z)$ such that* (1) *is satisfied everywhere on S.*

Proof. Denote the rings of S by $R_1, R_2, \cdots$. Without loss of generality we may assume that the r's are non-decreasing. Denote by l_i the last value of j such that $z=r_j$ is inside R_i. (This use of "inside" agrees with our previous use of the term if we take P to be the origin.) Let S_0 consist of those points of the rays $\alpha_1, \alpha_2, \cdots$ which are inside R_1. Let S_1 consist of the points of R_1 together with the points of the rays which are outside R_1 and inside R_2; and S_k the points of R_k together with those of the rays which are between R_k and R_{k+1}. Let $\epsilon_1, \epsilon_2, \cdots$ be a sequence of positive numbers such that $\sum \epsilon_i$ con-

verges to a given value ϵ. Let M_i be the upper bound of $G(z)$ in S_i. Let a_i be the radius of the smaller of the two circles bounding R_i.

Put $N_0 = M_0 + \epsilon$, $N_1 = M_1 + N_0 + \epsilon$. For brevity, we also denote by $\alpha_1, \alpha_2, \cdots$ the fractional multiplies of 2π that the rays $\alpha_1, \alpha_2, \cdots$ make with the axis of reals. We choose p_1 to be a positive integer such that $p_1\alpha_j$ is an integer for $j = 1, 2, \cdots, l_2$. Let

$$N_2 = \underset{z \text{ in } S_2}{\text{l.u.b.}} \left\{ M_2 + \epsilon + \left| N_0 + N_1\left(\frac{z}{a_1}\right)^{p_1} \right| \right\}.$$

We choose p_2 to be a positive integer such that (a) $p_2\alpha_j$ is an integer for $j = 1, 2, \cdots, l_3$, and (b) $|N_2(z/a_2)^{p_2}|$ is less than ϵ_2 inside and on R_1.

Continuing this process, at the typical stage we let

$$N_i = \underset{z \text{ in } S_i}{\text{l.u.b.}} \left\{ M_i + \epsilon + \left| N_0 + N_1\left(\frac{z}{a_i}\right)^{p_1} + \cdots + N_{i-1}\left(\frac{z}{a_{i-1}}\right)^{p_{i-1}} \right| \right\}.$$

Then we take p_i to be a positive integer so large that (a) $p_i\alpha_j$ is an integer for $j = 1, 2, \cdots, l_{i+1}$, and (b) $|N_i(z/a_i)^{p_i}|$ is less than ϵ_i inside and on R_{i-1}.

Now consider the function

$$f(z) = N_0 + \sum_{i=1}^{\infty} N_i\left(\frac{z}{a_i}\right)^{p_i}. \tag{3}$$

Let Σ be any finite closed region, and z a point of Σ. Then there is a q so large that

$$\left| N_i\left(\frac{z}{a_i}\right)^{p_i} \right| < \epsilon_i \qquad \text{for } z \text{ in } \Sigma \text{ and } i > q.$$

Since $\Sigma\epsilon_i$ converges, this means that (3) converges uniformly in Σ, and $f(z)$ is an integral function.

Moreover, if z is in one of the rings R_q, we can write

$$\begin{aligned} |f(z)| &\geqq \left| N_q\left(\frac{z}{a_q}\right)^{p_q} \right| - \left| N_0 + \sum_{i=1}^{q-1} N_i\left(\frac{z}{a_i}\right)^{p_i} \right| - \left| \sum_{i=q+1}^{\infty} N_i\left(\frac{z}{a_i}\right)^{p_i} \right| \\ &\geqq M_q \geqq G(z). \end{aligned}$$

If z is on one of the rays and in S_q, then

$$|f(z)| \geqq \sum_{i=q}^{\infty} N_i\left(\frac{z}{a_i}\right)^{p_i} - \left| N_0 + \sum_{i=1}^{q-1} N_i\left(\frac{z}{a_i}\right)^{p_i} \right| \geqq M_q \geqq G(z).$$

This concludes the proof of the theorem.

9. **An unsolved problem.** As mentioned in the Introduction, a problem, intermediate between Theorems IV and V, which the writer has been unable to answer, but whose solution could be used to advantage in the subsequent

part of this paper, is the following: Let $S_1, S_2, \cdots$ be a given infinite sequence of simply connected regions whose closures are non-intersecting and whose only sequential limit point is the point at infinity. Let $M_1, M_2, \cdots$ be a given sequence of positive numbers. Then, does there exist a *nonvanishing* integral function, or even a nonvanishing meromorphic function, $f(z)$, such that

$$|f(z)| \geqq M_i$$

for z in S_i? The available evidence leads to the conjecture that for every sequence of regions there will correspond M's for which no such integral function exists.

10. **A theorem on level curves.** The zeros of an analytic function are isolated. Conversely, according to Mittag-Leffler's theorem, given any isolated set I, there is a function which has a zero at each point of I and none elsewhere and which is analytic except at limit points of I.

The singularities of an analytic function form a closed set. For any given closed set B there exists an isolated set I, in the complement of B, whose derived set I' is the boundary of B. It follows that corresponding to every closed set B there is an analytic function whose singularities are precisely the points of B and whose zeros have every point of the boundary of B as a limit point.

The Mittag-Leffler theorem gives a complete characterization of the possible distribution of zeros of an analytic function, but it gives no information about the possible behavior of the function away from those zeros. The following theorem goes further, by preassigning not only the position of the zeros but something about the level curves as well.

THEOREM VII. *Let B be a given closed set and $I \subset C(B)$ be an isolated set such that I' is the boundary of B. Let $a_1, a_2, \cdots$ be the points of I, and $\gamma_1, \gamma_2, \cdots$ be a given sequence of mutually exterior circles, with the center of γ_i at a_i, such that no point of B is inside or on any γ_i. Then there exists a function $f(z)$ which is analytic except on B, which has a simple zero at each point of I, and whose level curve C: $|f(z)| = 1$ is such that the part of C which is inside γ_i is a Jordan curve separating a_i from γ_i, $i = 1, 2, \cdots$.*

Here "inside γ_i" is used in the sense of being on the same side of γ_i as a_i.

Proof. Let $h(z)$ be any function with simple zeros at $a_1, a_2, \cdots$ and no other zeros, and analytic except on B. Let $M_1, M_2, \cdots$ be the greatest lower bounds of $|h(z)|$ on $\gamma_1, \gamma_2, \cdots$, respectively. Let $g(z)$ be a function (whose existence is asserted by the corollary to Theorem IV) which is analytic except on B and which is such that $|g(z)| > 1/M_i$ on γ_i. Then the product $f(z) = h(z)g(z)$ will be analytic inside and on γ_i and $|f(z)| > 1$ on γ_i. Hence there will be a part of the level curve C which is inside γ_i and which separates a_i from γ_i. Thus $f(z)$ has the properties stated in the theorem.

Theorem VII is unsatisfactory in that it makes no assertion about the zeros of $f(z)$ other than those on I.

PART III

11. **Statement of the problem.** A known expansion theorem[6] asserts that for any set of n points $a_1, a_2, \cdots, a_n$ there exists a sequence of functions $\phi_s(z)$, $s=1, 2, \cdots$, satisfying the following two conditions:

(A) Each function $\phi_s(z)$ is analytic in a region including all the a's.

(B) Corresponding to any function $f(z)$ which is analytic at all the a's (but not necessarily analytic in any region containing more than one of the a's), there is a sequence of constants $c_1, c_2, \cdots$ such that

$$f(z) = \sum_{s=1}^{\infty} c_s\phi_s(z).$$

This series converges absolutely and uniformly in some neighborhood of each point a_i.

The case of chief interest is that in which $f(z)$ cannot be continued analytically from a_i to a_k, $i \neq k$. (For example, $f(z)$ might be e^z near a_1, $\sin z$ near a_2, z^2 near a_3, etc.)

We wish now to generalize this theorem by allowing n to become infinite, so that $a_1, a_2, \cdots$ will form an arbitrary isolated set I. In doing this, certain characteristic differences with the finite case arise:

(1) The set of limit points B of I will be non-vacuous. One would expect the functions $\phi_s(z)$ to be badly behaved on B, particularly if B divides the plane. We, therefore, replace condition (A) by the condition that $\phi_s(z)$ be analytic everywhere except on B.

(2) One might also expect that for a fixed set of functions $\{\phi_s(z)\}$ any corresponding $f(z)$ would of necessity satisfy certain uniformity conditions with respect to the points a_i. Such uniformity conditions are embodied in the following:

DEFINITION. *A function $f(z)$ will be said to belong to the class $G\{\theta_i, a_i\}$, where $\theta_1, \theta_2, \cdots$ is a sequence of positive numbers, provided $f(z)$ is analytic and uniformly bounded in a set of closed circular regions with centers at $a_1, a_2, \cdots$ and radii $\delta_1, \delta_2, \cdots$, where for some λ*

$$\delta_i\theta_i \geqq \lambda > 0, \qquad i = 1, 2, \cdots.$$

We propose to prove the following theorem:

THEOREM VIII. *Let $\{a_i\}$ be any isolated set of points with derived set B, and let $\{\theta_i\}$ be any sequence of positive numbers. Then there is a sequence of functions $\{\phi_s(z)\}$, each of which is analytic everywhere except on B, such that for any function $f(z)$ of class $G\{\theta_i, a_i\}$ there are numbers $c_1, c_2, \cdots$ for which*

$$f(z) = \sum_{s=1}^{\infty} c_s\phi_s(z).$$

---

(6) Tôhoku Mathematical Journal, vol. 43 (1937), pp. 246–251.

The series will converge absolutely and uniformly in some point set which contains a neighborhood of each point a_i.

For the above case in which n is finite, the functions $\{\phi_s(z)\}$ depend only on the points $a_1, a_2, \cdots, a_n$, and not at all on $f(z)$. In Theorem VIII the functions $\{\phi_s(z)\}$ depend on $\theta_1, \theta_2, \cdots$ as well as upon the points $\{a_i\}$; but they are otherwise independent of $f(z)$.

The proof of Theorem VIII will be carried out in steps. It will be shown first that there exists a particular sequence of points $a_1, a_2, \cdots$ for which the theorem holds. A transformation will then be performed which will send this particular set of a's into an arbitrary sequence.

12. **Proof of the expansion theorem for a particular sequence of points.** The particular set of points $a_1, a_2, \cdots$ to be considered in this section will be a certain set of integers $k_1, k_2, \cdots$, where $k_1=0$ and $k_{i+1}>k_i$. It is desired to expand a function $f(z)$ of class $G\{\theta_i, k_i\}$. Let $\gamma_1, \gamma_2, \cdots$ be a sequence of mutually exterior circles with centers at $k_1, k_2, \cdots$; with radii $\beta_1, \beta_2, \cdots$, where the β's have a finite upper bound β, and $\beta_i \leqq 1/\theta_i$; and in which $f(z)$ is analytic and uniformly bounded. According to Theorem VII there exists an integral function $\psi(z)$ which has a simple zero at each point k_i and whose level curve C_η: $|\psi(z)| = \eta$, $\eta \leqq 1$, is such that the part of C_η which is inside γ_i is a Jordan curve separating k_i from γ_i. Denote by $C_\eta^{(i)}$ this Jordan curve inside γ_i, and by $S_\eta^{(i)}$ the closed region containing k_i which is bounded by $C_\eta^{(i)}$. Let z_1 be any point in $S_\eta^{(1)}$, and z_i the point (there is one and just one) in $S_\eta^{(i)}$ where $\psi(z_i) = \psi(z_1)$.

Consider the infinite system of linear equations

$$\sum_{j=1}^{\infty} g_j(z_i) h_j(z_1) = f(z_i), \qquad i = 1, 2, \cdots, \tag{4}$$

where the $g_j(z)$ are integral functions and the $h_j(z_1)$ are regarded as unknowns. We wish to define the $g_j(z)$ so that the determinant of the equations will be normal and different from zero(7). Let

$$g_j(z) = e^{-\alpha_j(z-k_j)^2},$$

where $\alpha_1, \alpha_2, \cdots$ are positive numbers such that $\sum \alpha_j$ converges. For the determinant $\Delta = |g_j(z_i)|$ to be normal it is sufficient that (a) the product of the diagonal terms converge absolutely and (b) the sum of the non-diagonal terms converge absolutely.

For (a) the product of the diagonal terms is

$$\prod_{j=1}^{\infty} g_j(z_j) = \exp\left\{\sum_{j=1}^{\infty} - \alpha_j(z_j - k_j)^2\right\}.$$

(7) For a discussion of normal determinants with applications to infinite systems of linear equations see F. Riesz, *Les Systèmes d'Équations Linéaires*, 1913, §§20–30.

Since $|z_j-k_j|\leqq\beta_j\leqq\beta$, this product converges absolutely. The convergence is, moreover, uniform with respect to z_1 in $S_\eta^{(1)}$ and uniform for all sets of integers $\{k_j\}$. The value of the product is uniformly bounded away from zero for all z_1 in $S_\eta^{(1)}$ and all sequences of integers $\{k_j\}$.

For (b) we have the double series

$$\sum_{i,j=1}^{\infty}{}' g_j(z_i) = \sum_{i,j=1}^{\infty}{}' \exp\{-\alpha_j(k_i - k_j + \lambda_i)^2\}, \tag{5}$$

where we have put $z_i=k_i+\lambda_i$, $|\lambda_i|\leqq\beta_i\leqq\beta$. The prime indicates omission of terms for $i=j$.

Let ϵ be any positive number, and let $\{M_{ij}\}$, $i\neq j$, be any set of positive numbers such that $\sum M_{ij}=\epsilon$. We wish to show that the k's can be chosen so that the absolute value of the ijth term of (5) will be less than M_{ij}.

We have

$$|g_j(z_i)| < \exp\{-\alpha_j[(k_i - k_j)^2 - 2\beta|k_i - k_j| - \beta^2]\} \equiv K_{ij}.$$

The k's will now be chosen successively as follows: Let k_2 be so large that

$$\max(K_{12}, K_{21}) < \min(M_{12}, M_{21}).$$

Then k_3 can be chosen so that

$$\max_{j=1,2}(K_{j3}, K_{3j}) < \min_{j=1,2}(M_{j3}, M_{3j}).$$

In general, having already chosen $k_2, k_3, \cdots, k_{s-1}$, we then make k_s so large that

$$\max_{j<s}(K_{js}, K_{sj}) < \min_{j<s}(M_{js}, M_{sj}).$$

Hence the terms of the series (5) are dominated by the numbers M_{ij}, where the latter may be preassigned arbitrarily. By proper choice of the k's the series will converge absolutely; hence Δ is normal. Moreover, the k's can be chosen so that the value of Δ differs from the value of the diagonal term by less than ϵ. Hence, for suitable k's, $\Delta\neq 0$. From this point on, we assume that the k's are such that Δ is normal and nonvanishing.

Since $f(z)$ is bounded in the regions $S_\eta^{(i)}$ by a constant independent of i, the system of equations (4) will have one and just one solution $\{h_s(z_1)\}$ such that $|h_s(z_1)|$ is bounded for z_1 in $S_\eta^{(1)}$ by a constant independent of s. Since the convergence of the determinant is uniform, and since $f(z)$ is analytic in $S_\eta^{(1)}, S_\eta^{(2)}, \cdots$, then the functions $h_s(z_1)$ will each be analytic in $S_\eta^{(1)}$. Put

$$h_s(z_j) = h_s(z_1).$$

Then $h_s(z)$ will be defined and analytic in each region $S_\eta^{(i)}$. Hence

$$f(z) = \sum_{j=1}^{\infty} e^{-\alpha_j(z-k_j)^2} h_j(z)$$

for every z in $S_\eta^{(1)}, S_\eta^{(2)}, \cdots$. The series on the right converges absolutely and uniformly on the same set of values of z.

Since $h_j(z)$ is analytic in $S_\eta^{(1)}$, it may be expanded in an absolutely and uniformly convergent series of powers of $\psi(z)$:

$$h_j(z) = \sum_{m=0}^{\infty} c_{jm}[\psi(z)]^m.$$

But, since $h_j(z)$ and each term of its expansion is unchanged if z_1 is replaced by z_j, the expansion is equally valid for z in $S_\eta^{(2)}, S_\eta^{(3)}, \cdots$. Hence

$$f(z) = \sum_{j=1}^{\infty} e^{-\alpha_j(z-k_j)^2} \sum_{m=0}^{\infty} c_{jm}[\psi(z)]^m$$

for z in $S_\eta^{(1)}, S_\eta^{(2)}, \cdots$.

Let M be the upper bound of $h_j(z)$ in $S_\eta^{(1)}$. Then, by an analogue of Cauchy's inequalities,

$$| c_{jm} | \leqq M/\eta^m.$$

Hence, if z is in $S_{\eta'}^{(i)}$, $\eta' < \eta$,

$$\sum_{m=0}^{\infty} | c_{jm}[\psi(z)]^m | \leqq \sum_{m=0}^{\infty} \frac{M}{\eta^m} \eta'^m = \frac{M}{1 - \eta'/\eta}.$$

We have already shown that $\sum g_j(z)$ converges absolutely. It follows that the double series in

$$f(z) = \sum_{j=1, m=0}^{\infty} c_{jm} e^{-\alpha_j(z-k_j)^2} [\psi(z)]^m$$

converges absolutely and uniformly to $f(z)$ in the regions $\{S_{\eta'}^{(i)}\}$. Finally, $f(z)$ can be represented by any simple series that can be formed by rearrangement of the terms of this double series. This completes the proof of Theorem VIII in the special case of a particular function $f(z)$ of class $G\{\theta_i, k_i\}$. If $\bar{f}(z)$ is any other function of this same class, there will be a positive number $\lambda \leqq 1$ such that

$$\delta_i \geqq \lambda/\theta_i \geqq \lambda\beta_i.$$

Since $\psi(z)$ has a simple zero at $z = k_i$, it follows from Schwarz' lemma on lower bounds that $C_{\lambda\eta}^{(i)}$ will be inside the circle $|z - k_i| = \lambda\beta_i$. Hence $\bar{f}(z)$ will be analytic and uniformly bounded in the regions $S_{\lambda\eta}^{(i)}$ and will have an expansion in terms of the same set of functions as $f(z)$.

13. **Proof of the expansion theorem for an arbitrary isolated set.** Consider any isolated set of points $\{a_i\}$ in an extended complex plane of the variable x. There exists a function $T(x)=z$ which is analytic except on B and which has the values

$$T(a_i) = k_i, \qquad T'(a_i) = \sigma_i, \qquad i = 1, 2, \cdots,$$

where $\sigma_1, \sigma_2, \cdots$ is a given sequence of complex numbers whose absolute values have a positive lower bound σ.

To construct such a function, we may first use the Weierstrass factor theorem to build a function $T_1(x)$ with a second order zero at each point $a_1, a_2, \cdots$. Let

$$T_1(x) = (x - a_i)^2[\alpha_{i1} + \alpha_{i2}(x - a_i) + \cdots],$$

where $\alpha_{i1} \neq 0$. Next, by the Mittag-Leffler theorem, we form a function $T_2(x)$ with principal part at a_i as follows:

$$\frac{k_i}{\alpha_{i1}} \frac{1}{(x - a_i)^2} + \frac{\sigma_i\alpha_{i1} - k_i\alpha_{i2}}{\alpha_{i1}^2} \frac{1}{x - a_i}.$$

Then the product $T(x)=T_1(x)T_2(x)$ will have the desired properties.

There exists an inverse function $x=T^{-1}(z)$ which is analytic at each point k_i, and

$$T^{-1}(k_i) = a_i, \qquad T^{-1\prime}(k_i) = 1/\sigma_i.$$

Hence, for every ϵ, $0<\epsilon<\sigma$, there exist positive numbers π_i so small that in the regions $|z-k_i| \leqq \pi_i$, $T^{-1}(z)$ is analytic and

$$| T^{-1}(z) - T^{-1}(k_i) | \leqq | z - k_i | /(\sigma - \epsilon).$$

We suppose that $f(x)$ is of class $G\{\tau_i, a_i\}$. Then $F(z) \equiv f(T^{-1}(z))$ will be analytic and uniformly bounded if

$$| z - k_i | \leqq \lambda(\sigma - \epsilon)/\tau_i \quad \text{and} \quad \leqq \pi_i.$$

Hence, $F(z)$ will belong to the class $G\{\theta_i, k_i\}$ where $\theta_i = \max(\tau_i, 1/\pi_i)$.

By the special case of Theorem VIII in the previous section, there exists a sequence of integral functions $\phi_\nu(z)$ such that any function of class $G\{\theta_i, k_i\}$ can be expanded in the form $F(z)=\sum c_\nu\phi_\nu(z)$ for some set of neighborhoods of $k_1, k_2, \cdots$. Hence $f(x)=\sum c_\nu\phi_\nu(T(x))$ for some set of neighborhoods of $a_1, a_2, \cdots$. The functions $\phi_\nu(T(z))$ will be analytic except for limit points of the a's and will depend only on the class $G\{\tau_i, a_i\}$ to which $f(x)$ belongs. This completes the proof of Theorem VIII.

UNIVERSITY OF ILLINOIS,
URBANA, ILL.

ON THE DEGREE OF POLYNOMIAL APPROXIMATION TO ANALYTIC FUNCTIONS: PROBLEM β

BY

J. L. WALSH AND W. E. SEWELL

1. Introduction. Given a closed bounded point set C of the z-plane whose complement K is connected and possesses a Green's function $G(x, y)$ with pole at infinity; denote generically by C_ρ the locus $G(x, y) = \log \rho$, $1 < \rho$, in K. By Problem β we understand the following problem: If a function $f(z)$ is assumed analytic interior to a particular C_ρ, and possesses given continuity properties on or in the neighborhood of C_ρ, to study the degree of approximation by polynomials to $f(z)$ on C in the sense of Tchebycheff.

This problem has reached a fairly satisfactory solution in case $f(z)$ has generalized derivatives of various orders on C_ρ(1), and in case $f(z)$ is continuous on and within C_ρ, and its pth derivative satisfies a Lipschitz condition of order α on C_ρ(2); in the latter case, if C is bounded by a finite number of smooth mutually exterior Jordan curves, it follows (loc. cit.) that polynomials $p_n(z)$ of respective degrees n exist such that

$$|f(z) - p_n(z)| \leq M/\rho^n \cdot n^{p+\alpha}, \qquad z \text{ on } C, \tag{1}$$

where M is a constant depending on C and ρ but independent of n and z.

However, if $f(z)$ is not assumed continuous on C_ρ but merely to become infinite (if at all) sufficiently slowly, a result closely analogous to (1) exists:

$$|f(z) - p_n(z)| \leq M n^{p+\alpha}/\rho^n, \qquad z \text{ on } C, \tag{2}$$

where $p+\alpha$ is again positive and is a measure of the rapidity with which $f(z)$ becomes infinite. Such a result has already been considered by S. Bernstein [1926] and de la Vallée Poussin [1919] for the case that C is a segment of the axis of reals, and *provided $f(z)$ has only isolated singularities on C_ρ*. The primary object of the present paper is to establish (2) for more general point sets C (especially when C is the closed interior of an analytic Jordan curve) and for functions $f(z)$ not required to have only isolated singularities on C_ρ.

To be more explicit, we define (§2) a hierarchy of functions, thanks to certain theorems due to Hardy and Littlewood [1932], which includes both functions whose derivatives satisfy Lipschitz conditions of various orders and functions satisfying asymptotic inequalities in the neighborhood of C_ρ. This classification of functions is highly appropriate for our present discussion, for

Presented to the Society, September 12, 1940; received by the editors June 4, 1940.

(1) Sewell [1937]; numbers in brackets refer to the bibliography at the end of this paper.

(2) Walsh and Sewell [1940].

it is primarily based on (i) Lipschitz or asymptotic conditions for the functions, but is so constituted that (ii) integrals and derivatives of functions of a class belong automatically in specified new classes, likewise defined in terms of Lipschitz or asymptotic conditions; also (iii) each class implies a specific degree of approximation (Problem β), and conversely (iv) certain definite degrees of approximation imply that the function belongs to a uniquely determined class; in each case under (i) to (iv) the results are in a sense the best possible.

We study these questions of approximation (§§3, 4, 5) for the unit circle, (§8) for the line segment, annulus, and real axis, and (§7) for point sets which are the closed interiors of analytic Jordan curves. In §6 we consider approximation to functions with isolated singularities. We indicate (§9) the method of extending the above results on Tchebycheff approximation to approximation measured by a line integral. In §10 we consider the relation between integrated Lipschitz conditions and integral asymptotic conditions on the one hand and degree of approximation on the other hand. Finally in §11 we present more immediate but less thoroughgoing methods for obtaining portions of our results.

The methods and results here set forth have application to the study of approximation of harmonic functions by harmonic polynomials, an application which the writers plan to make on another occasion.

Henceforth in the present paper the degree of a polynomial is indicated consistently by its subscript; moreover the letter M with or without subscripts when used in an inequality of type (1) or (2) shall always represent a constant which may vary from inequality to inequality and depends on C and ρ but which is always independent of n and z.

2. **A classification of functions.** In the present section, the unit circle $|z|=1$ is denoted by γ. If the function $f(z)$ is analytic interior to γ, continuous in the corresponding closed region, and if $f^{(p)}(z)$, where $p\geqq 0$ is an integer, satisfies a Lipschitz condition on γ of order α, $0<\alpha\leqq 1$, we say that $f(z)$ is *of class* $L(p, \alpha)$ *on* γ. It is immaterial here whether we require that $f^{(p)}(z)$ and the Lipschitz condition should be one-dimensional or two-dimensional; compare Hardy and Littlewood [1932], Walsh and Sewell [1940]. It obviously follows that if $f(z)$ is of class $L(p, \alpha)$ on γ then the indefinite integral of $f(z)$ is of class $L(p+1, \alpha)$ on γ and (provided $p>0$) the derivative $f'(z)$ is of class $L(p-1, \alpha)$ on γ. In this connection it is appropriate to consider the following theorem due to Hardy and Littlewood [1932]:

THEOREM 2.1. *A necessary and sufficient condition that $f(z)$, analytic for $|z|<1$, should belong to class $L(0, \alpha)$, $0<\alpha\leqq 1$, is that*

$$|f'(re^{i\theta})| \leqq M(1-r)^{\alpha-1}, \qquad r<1, \tag{2.1}$$

where $z=re^{i\theta}$ and where M is independent of r and θ.

This theorem suggests a new definition: If the function $f(z)$ is analytic for $|z|<1$ and if we have

$$|f(re^{i\theta})| \leqq M(1-r)^{\alpha+p}, \qquad r<1,\ 0<\alpha\leqq 1, \tag{2.2}$$

where $p<0$ is an integer, where $z=re^{i\theta}$, and where M is independent of r and θ, then $f(z)$ is said to be *of class $L(p, \alpha)$ on γ*. With this terminology we prove

THEOREM 2.2. *If the function $f(z)$ is of class $L(p, \alpha)$ on γ, $0<\alpha\leqq 1$, then the indefinite integral of $f(z)$ is of class $L(p+1, \alpha)$ on γ unless $\alpha+p=-1$, and the derivative $f'(z)$ is of class $L(p-1, \alpha)$ on γ.*

We set

$$F(z) = \int_0^z f(z)dz; \tag{2.3}$$

our conclusion concerning $F(z)$ for $p\geqq 0$ has already been mentioned, and for $p=-1$ follows from Theorem 2.1. For $p<-1$ we take the path of integration in (2.3) a radius, which involves no loss of generality:

$$F(re^{i\theta}) = \int_0^r f(re^{i\theta})dr,$$

where θ is fixed. We have by (2.2)

$$|F(re^{i\theta})| \leqq M\int_0^r (1-r)^{\alpha+p}dr \leqq M'[(1-r)^{\alpha+(p+1)}-1], \tag{2.4}$$

from which our conclusion on $F(z)$ (and on any indefinite integral of $f(z)$) follows.

In the case $p>0$ the conclusion of Theorem 2.2 concerning $f'(z)$ has already been mentioned, and this conclusion for $p=0$ follows from Theorem 2.1. Suppose now $p<0$, so that $p+\alpha\leqq 0$. Let z be fixed interior to γ. We choose $\rho=\frac{1}{2}(1-|z|)$ and study the integral

$$f'(z) = \frac{1}{2\pi i}\int_{|t-z|=\rho} \frac{f(t)}{(t-z)^2}\,dt. \tag{2.5}$$

On the path of integration we have (2.2) satisfied, whence

$$|f'(z)| \leqq \frac{2M[1-|z|-\rho]^{\alpha+p}}{1-|z|} \leqq M_1(1-|z|)^{\alpha+p-1}, \tag{2.6}$$

as we were to prove. Theorem 2.2 is established.

It will be noticed that the proof of (2.4) fails in the case $\alpha+p=-1$, that is to say, in the case $p=-2$, $\alpha=1$. In this connection it is useful to introduce a new definition, namely that $f(z)$ *shall be of class $L'(p, 1)$, $p\geqq -1$, provided $f(z)$ is analytic interior to γ, and provided $f^{(p+2)}(z)$ is of class $L(-2, 1)$.*

We make the following observation:

THEOREM 2.3. *If $f(z)$ is analytic and uniformly bounded interior to γ, then $f(z)$ is of class $L'(-1, 1)$.*

As above we use equation (2.5), where $\rho=(1-|z|)/2$; then we have

$$|f'(z)| \leqq \frac{M}{2\pi}\frac{2\pi\rho}{\rho^2} = 2M(1-|z|)^{-1};$$

thus $f'(z)$ is of class $L(-2, 1)$, so the theorem follows.

We shall now establish

THEOREM 2.4. *If $f(z)$ is of class $L'(p, 1)$, $p>-1$, then $f'(z)$ is of class $L'(p-1, 1)$; moreover $f^{(p+2+k)}(z)$, where k is a positive integer, is of class $L(-1-k, 1)$.*

Also, if $f(z)$ is of class $L'(p, 1)$ then for r near unity we have

$$|f^{(p+1)}(re^{i\theta})| \leqq M|\log(1-r)|, \tag{2.7}$$

and on the radii $f^{(p)}(z)$ satisfies the pseudo-Lipschitz condition for r near unity

$$|f^{(p)}(e^{i\theta}) - f^{(p)}(re^{i\theta})| \leqq M'(1-r)|\log(1-r)|, \tag{2.8}$$

where M' is independent of θ; under these conditions the qth integral of $f(z)$ is of class $L'(p+q, 1)$, $q>0$.

If $f(z)$ is of class $L'(p, 1)$, with $p>-1$, we have by definition $|f^{(p+2)}(z)| \leqq M(1-r)^{-1}$; but $f^{(p+2)}(z)$ is the derivative of order $p+1$ of $f'(z)$ and hence, also by the definition of class $L'(p, 1)$, the function $f'(z)$ is of class $L'(p-1, 1)$. Furthermore it is clear from the proof of (2.6) that $|f^{(p+2+k)}(z)| \leqq M(1-r)^{-1-k}$, k a positive integer, and hence $f^{(p+2+k)}(z)$ is of class $L(-2-k, 1)$ by definition.

If $f(z)$ is of class $L'(p, 1)$, an inequality on $f^{(p+1)}(re^{i\theta})$ follows directly from the inequality on $f^{(p+2)}(re^{i\theta})$:

$$|f^{(p+1)}(re^{i\theta}) - f^{(p+1)}(0)| = \left|\int_0^r f^{(p+2)}(re^{i\theta})dr\right|$$

$$\leqq M\left|\int_0^r \frac{dr}{1-r}\right| = M|\log(1-r)|,$$

and since $f^{(p+1)}(0)$ is a constant we have the inequality of the theorem. The function $f^{(p)}(z)$ can be defined on the boundary as the integral of its derivative and the pseudo-Lipschitz condition (2.8) is an immediate consequence of the integration of (2.7) from r to 1 along an arbitrary radius. The remark about the qth integral follows from the definition and the fact that the derivative of an indefinite integral is the function itself under the above conditions.

The uniform pseudo-Lipschitz condition (2.8) on the radii, for functions of class $L'(0, 1)$, implies a similar condition on the circumference:

COROLLARY. *If $f(z)$ is of class $L'(0, 1)$, then $f(z)$ is continuous in the two-dimensional sense on $|z|=1$, and satisfies on that circumference a uniform pseudo-Lipschitz condition of the form*

$$|f(e^{i\theta}) - f(e^{i\theta'})| \leqq M_1 |\log|\theta - \theta'|| \cdot |\theta - \theta'|, \tag{2.9}$$

where $|\theta-\theta'|$ is sufficiently small.

Let θ and θ' be given, $|\theta-\theta'|<1$. By (2.8) we have

$$|f(e^{i\theta}) - f(re^{i\theta})| \leqq M'(1-r)|\log(1-r)|,$$

$$|f(e^{i\theta'}) - f(re^{i\theta'})| \leqq M'(1-r)|\log(1-r)|.$$

Also by (2.7) we have $|f'(re^{i\theta})| \leqq M|\log(1-r)|$,

$$|f(re^{i\theta}) - f(re^{i\theta'})| \leqq M|\log(1-r)| \cdot |\theta - \theta'|.$$

The choice $1-r=|\theta-\theta'|$ now yields (2.9).

3. **Degree of approximation, unit circle.** We now present a proof of the following theorem, which connects the class $L(p, \alpha)$ whether p is positive, negative, or zero with degree of approximation:

THEOREM 3.1. *If $f(z)$ belongs to class $L(p, \alpha)$, $0<\alpha\leqq 1$, on γ: $|z|=1$, then there exist polynomials $p_n(z)$ such that we have on the circle $|z|=1/\rho<1$*

$$|f(z) - p_n(z)| \leqq M/\rho^n \cdot n^{p+\alpha}. \tag{3.1}$$

For the case $p\geqq 0$, Theorem 3.1 has already been established [Walsh and Sewell, 1940]; a new proof is given below, Theorem 10.5, second proof. For the case $p<0$ we proceed as follows. The formula

$$f(z) - \sum_{m=0}^{n} a_m z^m = \frac{1}{2\pi i}\int_{|t|=r<1} \frac{z^{n+1}f(t)}{t^{n+1}(t-z)}\,dt, \qquad |z|<r,$$

where $f(z)=\sum_{m=0}^{\infty} a_m z^m$, $|z|<1$, is well known. Thus we obtain

$$\left| f(z) - \sum_{m=0}^{n} a_m z^m \right| \leqq \frac{M}{2\pi}\left(\frac{1}{\rho}\right)^{n+1} \frac{(1-r)^{p+\alpha}2\pi r}{r^{n+1}(r-1/\rho)}, \qquad |z| \leqq 1/\rho < r.$$

If we let $r_n=1-1/n$, we have for n sufficiently large

$$\left| f(z) - \sum_{m=0}^{n} a_m z^m \right| \leqq \frac{M_1(1/n)^{p+\alpha}}{\rho^{n+1}(1-1/n)^{n+1}} < \frac{M_2}{\rho^n n^{p+\alpha}}, \qquad |z| \leqq 1/\rho,$$

since $(1-1/n)^n$ approaches $1/e$ as n becomes infinite. For a suitably chosen constant M_2 this inequality is valid for all n, $n=1, 2, \cdots$, and the proof of the theorem is complete.

By way of complement to Theorem 3.1 we state the following theorem, whose proof is postponed until §4:

THEOREM 3.2. *If $f(z)$ belongs to class $L'(p, 1)$, $p \geqq -1$, then there exist polynomials $p_n(z)$ such that we have on the circle $|z| = 1/\rho < 1$*

$$|f(z) - p_n(z)| \leqq M/\rho^n \cdot n^{p+1}. \tag{3.2}$$

Of course the hypothesis of Theorem 3.2 is less restrictive than that of Theorem 3.1 in the case $\alpha = 1$, $p \geqq -1$.

4. **Operations with approximating sequences.** It is our object in the present section to show how certain assumptions on a function imply immediate results on degree of approximation by polynomials to the various derivatives and integrals of that function.

THEOREM 4.1. *Let $f'(z)$ be of class $L(p, \alpha)$, $p \leqq -1$, $0 < \alpha \leqq 1$. Let $p_n'(z)$ denote the sum of the first $n+1$ terms of the Taylor development of $f'(z)$. Then we have for $|z| = 1/\rho < 1$*

$$\left| \int_0^z [f'(z) - p_n'(z)]dz \right| \leqq M/\rho^n \cdot n^{p+\alpha+1}.$$

We have the usual formula

$$f'(z) - p_n'(z) = \frac{1}{2\pi i} \int_{|t|=r} \frac{z^{n+1} f'(t)}{t^{n+1}(t-z)} dt, \qquad |z| \leqq 1/\rho < r < 1,$$

and hence

$$\int_0^z [f'(z) - p_n'(z)]dz = \frac{1}{2\pi i} \int_{|t|=r} \frac{f'(t)}{t^{n+1}} dt \int_0^z \frac{z^{n+1}}{t-z} dz,$$

where for simplicity the path of integration is chosen along a radius. But for $|t| > |z|$ we have

$$\int_0^z \frac{z^{n+1}}{t-z} dz = \frac{1}{t}\left[\frac{z^{n+2}}{n+2} + \frac{z^{n+3}}{(n+3)t} + \cdots\right];$$

for $|t| = r$ and $|z| \leqq 1/\rho$ the modulus of this function is dominated by

$$\frac{1}{n\rho^n}\left[\frac{1}{1 - 1/r\rho}\right].$$

Thus by the method employed in the proof of Theorem 3.1 with $r_n = 1 - 1/n$, we obtain the inequality

$$\left| \int_0^z [f'(z) - p_n'(z)]dz \right| \leqq M_1/\rho^n n^{p+\alpha+1}, \qquad |z| \leqq 1/\rho,$$

and the proof of the theorem is complete.

Theorem 4.1 is stated merely for the first integral of a function of class

$L(p, \alpha)$, but obviously extends to the iterated indefinite integrals of every order. Theorem 4.1 thus yields a new proof of Theorem 3.1 for the case $p \geqq 0$, and furnishes a proof of Theorem 3.2, which was not proved previously.

Another theorem relating to integration of approximating sequences (and which extends to iterated integrals of arbitrary order) is

THEOREM 4.2. *Let $f(z)$ be analytic interior to γ and continuous on γ. Let there exist polynomials $P_n(z)$ such that we have*

$$|f(z) - P_n(z)| \leqq \epsilon_n, \qquad z \text{ on } \gamma.$$

Let $p_n(z)$ denote the sum of the first $n+1$ terms of the Taylor development of $f(z)$. Then for $|z| = 1/\rho < 1$ we have

$$\left| \int_0^z [f(z) - p_n(z)]dz \right| \leqq M\epsilon_n/n\cdot\rho^n.$$

Theorem 4.2 admits of a relatively simple proof, but is to be reconsidered later (§7), and hence is not established in detail here. It may be noted that Theorem 4.2 with its extension to higher integrals yields by a transformation $z' = \sigma z$ a new proof of Theorem 3.1 for the case $p > 0$ by virtue of Theorem 3.1 itself for the case $p = 0$.

In connection with the differentiation of approximating sequences we also have two results analogous to Theorems 4.1 and 4.2:

THEOREM 4.3. *Let $f(z)$ be of class $L(p, \alpha)$, $p \leqq -1$, $0 < \alpha \leqq 1$. Let $p_n(z)$ denote the sum of the first $n+1$ terms of the Taylor development of $f(z)$. Then we have for $|z| = 1/\rho < 1$*

$$|f'(z) - p_n'(z)| \leqq M/\rho^n\cdot n^{p+\alpha-1}.$$

Theorem 4.3 can be proved by the method used for Theorem 4.1, and is in a sense to be generalized later as well (Theorem 7.9).

THEOREM 4.4. *Let $f(z)$ be analytic interior to γ: $|z| < 1$, and continuous on γ. Let there exist polynomials $P_n(z)$ such that we have*

$$|f(z) - P_n(z)| \leqq \epsilon_n, \qquad z \text{ on } \gamma.$$

Let $p_n(z)$ denote the sum of the first $n+1$ terms of the Taylor development of $f(z)$. Then for $|z| = 1/\rho < 1$ we have

$$|f'(z) - p_n'(z)| \leqq Mn\epsilon_n/\rho^n.$$

The proof of Theorem 4.4 is likewise postponed (compare Theorem 7.10 below). Both Theorem 4.3 and Theorem 4.4 extend at once to higher derivatives.

5. **Inverse problem. Examples.** In the direction of a converse to Theorem 3.1 we establish

THEOREM 5.1. *Let there exist polynomials $p_n(z)$ such that we have*

$$|f(z) - p_n(z)| \leqq M/\rho^n \cdot n^{p+\alpha+1}, \qquad |z| = 1/\rho < 1, \tag{5.1}$$

where p is integral and $0<\alpha\leqq 1$. Then $f(z)$, if properly extended analytically from the circle $|z|=1/\rho$, belongs to class $L(p, \alpha)$ on $|z|=1$ if $p+\alpha+1$ is not a positive integer, and to class $L'(p, \alpha)$ if $p+\alpha+1$ is a positive integer.

Theorem 5.1 has already been established for the case $p\geqq 0$, $\alpha<1$ (Walsh and Sewell [1940]). From (5.1) we may now write

$$|f(z) - p_{n+1}(z)| \leqq M/\rho^{n+1}\cdot(n+1)^{p+\alpha+1}, \qquad |z| = 1/\rho,$$

whence also by (5.1) we have, whether $p+\alpha+1$ is positive or nonpositive,

$$|p_{n+1}(z) - p_n(z)| \leqq 2M_0/\rho^n \cdot n^{p+\alpha+1}, \qquad |z| = 1/\rho. \tag{5.2}$$

The extended Bernstein Lemma (e.g. Walsh [1935, p. 77]) then yields

$$|p_{n+1}(z) - p_n(z)| \leqq 2M_0\rho r^{n+1}/n^{p+\alpha+1}, \qquad |z| = r > 1/\rho. \tag{5.3}$$

We define $f(z)$ in the region $1/\rho<|z|<1$ by means of the convergent sequence $p_n(z)$, so from (5.1) we see that $f(z)$ is analytic throughout the region $|z|<1$. On the circle $|z|=r<1$, $r>1/\rho$, we can write

$$f(z) = p_1(z) + [p_2(z) - p_1(z)] + [p_3(z) - p_2(z)] + \cdots,$$

$$|f(z)| \leqq M_1\sum_{n=1}^{\infty} r^n/n^{p+\alpha+1}. \tag{5.4}$$

If $p+\alpha<0$ we write $q=p+\alpha+1$,

$$\sum_{n=2}^{\infty} r^n/n^q \leqq \int_0^\infty r^x x^{-q}dx = \int_0^\infty e^{x\log r}x^{-q}dx$$
$$= \Gamma(1-q)(-\log r)^{q-1} \leqq M_2(1-r)^{q-1};$$

thus we have for $|z|=r<1$

$$|f(z)| \leqq M(1-r)^{p+\alpha},$$

so the conclusion follows unless $p+\alpha+1$ is a positive integer.

If now $p+\alpha+1$ is a positive integer, we write from (5.1) by the least-square property of the Taylor development of $f(z)\equiv\sum_{k=0}^{\infty}a_kz^k$,

$$\frac{M^2}{\rho^{2n}n^{2(p+\alpha+1)}} \geqq \frac{\rho}{2\pi}\int_{|z|=1/\rho} |f(z) - p_n(z)|^2|dz|$$
$$\geqq \frac{\rho}{2\pi}\int_{|z|=1/\rho}\left|f(z) - \sum_{k=0}^{n} a_kz^k\right|^2|dz| = \sum_{n+1}^{\infty}|a_k|^2/\rho^{2k};$$

it follows that we have for every n

$$|a_n| \leqq M_1/n^{p+\alpha+1}, \qquad |n^{p+\alpha+1}a_n| \leqq M_1,$$

$$|(n+p+\alpha)(n+p+\alpha-1)\cdots(n+1)na_n| \leqq M_2.$$

As in the use of (5.4) for $p+\alpha<0$ it follows now that $f^{(p+\alpha+1)}(z)$ is of class $L(-2, 1)$, hence that $f(z)$ is of class $L'(p, \alpha)$; Theorem 5.1 is established.

It will be noticed that there is a discrepancy of unity in the exponents of n in Theorems 3.1 and 5.1, in such a way that those theorems are not exact converses of each other. *This discrepancy is inherent in the nature of the problem,* as we shall show by examples. Such examples have already been provided [Walsh and Sewell, 1940] for the case $p \geqq 0$; we consider now the case $p<0$.

Let $p<0$ be given, and also α, $0<\alpha<1$. If for every function of class $L(p, \alpha)$ we could establish the existence of polynomials $p_n(z)$ with

$$|f(z) - p_n(z)| \leqq \epsilon_n, \qquad |z| = 1/\rho < 1, \tag{5.5}$$

where

$$\lim_{n\to\infty} \rho^n \cdot n^{p+\alpha}\epsilon_n = 0, \tag{5.6}$$

we should have by virtue of the least-square property of the Taylor development

$$\frac{2\pi\epsilon_n^2}{\rho} \geqq \int_{|z|=1/\rho} |f(z) - p_n(z)|^2 \cdot |dz| \geqq \int_{|z|=1/\rho} |f(z) - s_n(z)|^2 \cdot |dz|,$$

where $s_n(z)$ is the sum of the first $n+1$ terms of the Taylor development $f(z) \equiv \sum_{n=0}^{\infty} a_n z^n$. Let $F^{(p)}(z)$ denote the $(-p)$th indefinite integral of $f(z)$, where the constants of integration at the origin are chosen to vanish:

$$F^{(p)}(z) = \sum_{n=0}^{\infty} \frac{a_n z^{n-p}}{(n+1)(n+2)\cdots(n-p)}.$$

Thus we have

$$\begin{aligned} \int_{|z|=1/\rho} \left| F^{(p)}(z) - \sum_{k=0}^{n} \frac{a_k z^{k-p}}{(k+1)(k+2)\cdots(k-p)} \right|^2 |dz| \\ = \frac{2\pi}{\rho} \sum_{k=n+1}^{\infty} \frac{|a_k|^2}{\rho^{2k-2p}(k+1)^2(k+2)^2\cdots(k-p)^2} \\ \leqq \frac{2\pi M}{\rho n^{-2p}} \sum_{k=n+1}^{\infty} \frac{|a_k|^2}{\rho^{2k}} \\ = \frac{M}{n^{-2p}} \int_{|z|=1/\rho} |f(z) - s_n(z)|^2 \cdot |dz| \leqq \frac{2\pi M \epsilon_n^2}{\rho n^{-2p}}. \end{aligned} \tag{5.7}$$

The extreme members of (5.7) form an inequality valid for an *arbitrary* function $F^{(p)}(z)$ of class $L(0, \alpha)$, an inequality which taken together with (5.6) has already been shown (loc. cit.) to be impossible.

The reasoning just given does not apply to the case $\alpha=1$, but for this case we can establish a less precise result. We shall show that it is not possible to prove for every function of class $L(p, 1)$ the existence of polynomials $p_n(z)$ such that we have (5.5) valid with

$$\epsilon_n \leqq M/\rho^n \cdot n^{p+1+\delta}, \qquad \delta > 0. \tag{5.8}$$

If (5.8) could be proved for every function $f(z)$ of class $L(p, 1)$, inequality (5.8) could be proved for every function of class $L(p+1, \delta_1)$, $0<\delta_1<\delta$, which is necessarily also of class $L(p, 1)$; but we have just proved that (5.5) and (5.8) cannot be established for all functions of the class $L(p+1, \delta_1)$; this remark completes our proof that Theorem 3.1 cannot be essentially improved, in the sense that for arbitrary p and α the exponent of n in the second member of (3.1) can be replaced by no larger number.

We show now that Theorem 5.1 cannot be improved, in the sense that in (5.1) the exponent of n in the second member can be replaced by no smaller number. Let p and α be given, $0<\alpha<1$. We choose the function

$$f(z) \equiv (1-z)^{p+\alpha} \equiv \sum a_n z^n,$$

from which there follows (e.g., de la Vallée Poussin [1914, §399])

$$|a_n| \leqq M/n^{p+\alpha+1}. \tag{5.9}$$

Thus we have

$$\left| f(z) - \sum_{k=0}^{n} a_k z^k \right| \leqq \sum_{k=n+1}^{\infty} M/\rho^k \cdot k^{p+\alpha+1}, \qquad |z| = 1/\rho.$$

Since $\rho^{k/2} \cdot k^{p+\alpha+1}$ increases with k, for k sufficiently large, we find for the last sum the bound

$$M\rho^{-n/2} \cdot n^{-p-\alpha-1} \sum_{k=n+1}^{\infty} \rho^{-k/2} = M_1/\rho^n \cdot n^{p+\alpha+1}.$$

That is to say, we have exhibited a function $f(z)$ of class $L(p, \alpha)$ *and of no higher class* for which (5.1) holds; thus for arbitrary p and $0<\alpha<1$ the exponent of n in (5.1) cannot be decreased without altering the conclusion of Theorem 5.1; this conclusion applies, by supplementary reasoning similar to that used in connection with (5.8), also for arbitrary p with $\alpha=1$.

6. **Degree of approximation—isolated singularities.** In the preceding section (§5) it was shown that if $f(z)=(1-z)^{p+\alpha}=\sum_{m=0}^{\infty} a_m z^m$ then

$$\left| f(z) - \sum_{m=0}^{n} a_m z^m \right| \leqq M/\rho^n \cdot n^{p+\alpha+1}, \qquad |z| \leqq 1/\rho, \tag{6.1}$$

a higher degree of approximation than is asserted in Theorem 3.1; for $f(z)$ is of class $L(p, \alpha)$, unless $p+\alpha$ is a non-negative integer, and is of no higher class. Functions with isolated singularities are thus of particular interest in the study of the degree of approximation; this section is devoted to an investigation of such functions.

We state a generalization of the above conclusion:

THEOREM 6.1. *Let*

$$\begin{aligned} f(z) &= F_1(z) + F_2(z) + \cdots + F_\sigma(z) \\ &\quad + k_1(z - z_1)^{h_1} + \cdots + k_\mu(z - z_\mu)^{h_\mu}, \qquad |z_j| = 1, j = 1, \cdots, \mu, \end{aligned}$$

where $F_i(z)$, $i=1, \cdots, \sigma$, is of class $L(p_i, \alpha_i)$ or $L'(p_i, \alpha_i)$. Let $H_i = p_i+\alpha_i$ and $h=\min(h_j-1, H_i)$. Then there exist polynomials $p_n(z)$ such that

$$|f(z) - p_n(z)| \leqq M/\rho^n \cdot n^h, \qquad |z| = 1/\rho.$$

The proof simply consists in applying Theorem 3.1 and the conclusion (6.1), and is left to the reader. If here $f(z) = \sum a_n z^n$, we have $|a_n| \leqq M_1/n^h$.

In a similar way we obtain for the special function $f(z) = \log(1-z)$, which is of class $L'(-1, 1)$, a stronger result than that of Theorem 3.2:

THEOREM 6.2. *Let $f(z) = \log(1-z) = \sum_{m=1}^\infty z^m/m$. Then we have*

$$\left| f(z) - \sum_{m=1}^n z^m/m \right| \leqq M/\rho^n \cdot n, \qquad |z| = 1/\rho < 1.$$

These theorems can be extended to somewhat more general functions by means of certain inequalities concerning multiplication of series. [See, e.g., Hardy and Littlewood, 1935.] If $f(z) = \sum_{m=0}^\infty a_m z^m$, $g(z) = \sum_{m=0}^\infty b_m z^m$, $f(z)g(z) = \sum_{m=0}^\infty c_m z^m$, and if $|a_m| \leqq M_1/r^m \cdot m^h$, $|b_m| \leqq M_2/\rho^m$, $1 \leqq r < \rho$, then $|c_m| \leqq M/r^m m^h$. Thus we have

THEOREM 6.3. *Under the hypothesis of Theorem 6.1 or 6.2, the conclusion is valid if the function $f(z)$ is replaced by the product $f(z)g(z)$, where $g(z)$ is analytic in $|z| \leqq 1$.*

We also have in the above notation as a consequence of the inequalities $|a_m| \leqq M_1/m^h$, $|b_m| \leqq M_2/m^l$, the following inequalities:

$$\begin{aligned} |c_m| &\leqq M_3/m^{h+l-1}, && 1 > h \geqq l, \\ |c_m| &\leqq M_3/m^h, && l \geqq h > 1,\ l > h \geqq 1, \\ |c_m| &\leqq M_3 \log m/m^{h+l-1}, && 1 = h \geqq l. \end{aligned}$$

Thus we have

THEOREM 6.4. *Let $f(z) = (z-z_1)^{h_1-1}(z-z_2)^{h_2-1} = \sum_{m=0}^\infty a_m z^m$, $|z_1| = |z_2| = 1$, $z_1 \neq z_2$. Then we have*

$$\left| f(z) - \sum_{m=0}^{n} a_m z^m \right| \leqq \epsilon_n, \qquad |z| = 1/\rho,$$

where

$$\epsilon_n = M/\rho^n \cdot n^{h_1+h_2-1}, \qquad 1 > h_1 \geqq h_2;$$

$$\epsilon_n = M/\rho^n \cdot n^{h_1}, \qquad h_2 \geqq h_1 > 1, \text{ or } h_2 > h_1 \geqq 1;$$

$$\epsilon_n = M \log n/\rho^n \cdot n^{h_1+h_2-1}, \qquad 1 = h_1 \geqq h_2.$$

The extension of Theorem 6.4 to functions of the type $\prod_{k=1}^{n}(z-z_k)^{h_k-1}$ is immediate; details can be supplied by the reader. Also, Theorems 6.1–6.4 extend with identical conclusions to approximation on an arbitrary analytic Jordan curve, by replacing the Taylor development by the Faber [1920] development of the function.

7. **Extensions to more general regions.** It is obvious that much of the preceding discussion can be applied to the study of approximation on point sets more general than circles; we proceed to discuss some of the details of this extension. Broadly considered, the extension (for instance Theorems 7.5 7.9) applies to Jordan curves which are required to be smooth but not necessarily analytic; however, some of the following methods of proof (Theorems 7.7 and 7.8) apply only to analytic Jordan curves, so for simplicity we restrict ourselves to that case.

The reader may notice that some of the following treatment (e.g., Theorems 7.5, 7.9, 7.10, 7.11) applies also to approximation on point sets which are not connected but are bounded by disjoint analytic Jordan curves, provided C_ρ has no multiple points.

Definition. *Let Γ be an analytic Jordan curve in the z-plane. Let the interior of Γ be mapped conformally onto the interior of γ: $|w|=1$, by the transformation $w=\Phi(z)$, $z=\Psi(w)$. The function $f(z)$ analytic interior to Γ is said to be of class $L(p, \alpha)$ on Γ if the function $f[\Psi(w)]$ (suitably defined on γ if necessary) is of class $L(p, \alpha)$ on γ, where $0<\alpha\leqq 1$ and p is an integer, positive, negative, or zero.*

Thanks to the analyticity of the Jordan curves considered, and of the consequent continuity of the derivatives of the mapping functions in the closed regions, the following theorems are immediate consequences of the discussion of §2:

Theorem 7.1. *If the function $f(z)$ is continuous on and within the analytic Jordan curve Γ, then a necessary and sufficient condition that $f(z)$ be of class $L(p, \alpha)$ on Γ with $p\geqq 0$ is that $f^{(p)}(z)$ satisfy on Γ a Lipschitz condition of order α.*

Theorem 7.2. *Let Γ be an analytic Jordan curve, and let $\Gamma(\rho)$ be a sequence of analytic Jordan curves interior to Γ defined for all values of ρ in an interval $\rho_0\leqq\rho<\rho_1$ by an equation of the form*

$$\Gamma(\rho):\quad |F(z)| = \rho, \tag{7.1}$$

where $F(z)$ is analytic on Γ, with $F'(z)$ different from zero on Γ, and with the property $|F(z)| = \rho_1$ on Γ. Then a necessary and sufficient condition that a function $f(z)$ be of class $L(p, \alpha)$ on Γ with $p<0$ is

$$|f(z)| \leqq N(\rho_1 - \rho)^{p+\alpha}, \qquad z \text{ on } \Gamma(\rho), \tag{7.2}$$

where N is independent of z and ρ.

The property expressed by (7.1) and (7.2) is independent of the particular analytic function $F(z)$ considered.

THEOREM 7.3. *If the function $f(z)$ is of class $L(p, \alpha)$ on the analytic Jordan curve Γ, $0<\alpha<1$, then the indefinite integral of $f(z)$ is of class $L(p+1, \alpha)$ on Γ, and the derivative $f'(z)$ is of class $L(p-1, \alpha)$ on Γ.*

The proof is easy and is left to the reader.

The class $L'(p, 1)$, $p \geqq -1$ on Γ is defined as the transform of the class $L'(p, 1)$ on γ. Analogous to Theorem 7.3 we have

THEOREM 7.4. *If the function $f(z)$ is analytic and bounded interior to Γ, it is of class $L'(-1, 1)$ on Γ.*

If $f(z)$ is of class $L'(p, 1)$, $p>-1$, on Γ, then $f'(z)$ is of class $L'(p-1, 1)$ on Γ; moreover $f^{(p+2+k)}(z)$, $k>0$, is of class $L(-2-k, 1)$ on Γ; the qth integral of $f(z)$ is of class $L'(p+q, 1)$.

Theorems 7.3 and 7.4 are the respective extensions of Theorems 2.2 and 2.3 together with 2.4. Likewise the study of degree of approximation for Γ can be treated precisely like the study for the unit circle (§3). We leave to the reader the proof of the extension of Theorem 3.1, already established [Walsh and Sewell, 1940] for $p \geqq 0$, and to which the method of Theorem 3.1 applies for $p<0$ with the interpolation formula for equidistributed points:

THEOREM 7.5. *Let C be an analytic Jordan curve, and let the function $f(z)$ analytic interior to C_ρ be of class $L(p, \alpha)$ on C_ρ. Then there exist polynomials $p_n(z)$ such that we have on C*

$$|f(z) - p_n(z)| \leqq M/\rho^n \cdot n^{p+\alpha}. \tag{7.3}$$

We shall indicate the proof of the extension of Theorem 3.2:

THEOREM 7.6. *Let C be an analytic Jordan curve, and let the function $f(z)$ analytic interior to C_ρ be of class $L'(p, 1)$ on C_ρ, $p \geqq -1$. Then there exist polynomials $p_n(z)$ such that we have (7.3) valid on C.*

Theorem 3.2 follows from Theorem 4.1 precisely as Theorem 7.6 follows from a general result of which Theorem 4.1 is a limiting case:

THEOREM 7.7(3). *Let C be an analytic Jordan curve and let $f(z)$ be analytic in the interior of C_ρ. Let $f'(z)$ be of class $L(p, \alpha)$, $p \leqq -1$, $0<\alpha\leqq 1$ on C_ρ, and let $p_n'(z)$ denote the polynomial of degree n defined by interpolation to $f'(z)$ in points uniformly distributed on a suitable level curve $C_{1-\delta}$ interior to C belonging to the analytic family of curves C_ν. Then we have for z on C*

$$\left|\int_a^z [f'(z) - p_n'(z)]dz\right| \leqq M/\rho^n n^{p+\alpha+1},$$

where a is an arbitrary point interior to $C_{1-\delta}$, and where the path of integration contains no point exterior to C.

Theorem 7.7 is stated merely for the first integral, but extends at once to an arbitrary integral.

We use the well known Lagrange-Hermite interpolation formula for z interior to C_r

$$f'(z) - p_n'(z) = \frac{1}{2\pi i}\int_{C_r} \frac{\omega_n(z)f'(t)}{\omega_n(t)(t - z)} dt, \qquad 1 < r < \rho,$$

where $\omega_n(z) \equiv (z-z_1)(z-z_2) \cdots (z-z_{n+1})$, the points z_j lying interior to the curve C_r. Then we have for z interior to C_r

$$\int_a^z [f'(z) - p_n'(z)]dz = \frac{1}{2\pi i}\int_{C_r} \frac{f'(t)dt}{\omega_n(t)}\int_a^z \frac{\omega_n(z)dz}{t - z}.$$

Let $w=\phi(z)$ map the exterior of C onto $|w|>1$ with $\phi(\infty)=\infty$, and let $\delta>0$ be chosen so small that the locus $C_{1-\delta}$: $|\phi(z)|=1-\delta$ is an analytic Jordan curve interior to C, with $\phi(z)$ analytic (except at infinity) and univalent throughout the closed exterior of $C_{1-\delta}$. We set $e^\sigma=(1-\delta)/|\phi'(\infty)|=(1-\delta)\Delta$.

We make use of the inequality [Curtiss, 1935 or Walsh and Sewell, 1940]

$$e^{-M} \leqq \left|\frac{\omega_n(z)}{e^{(n+1)\sigma}(w^{n+1} - 1)}\right| \leqq e^M, \quad z \text{ on or exterior to } C_{1-\delta},$$

where the points z_j are chosen as equally distributed points on $C_{1-\delta}$, and where w now and henceforth represents the function $w=\phi(z)/(1-\delta)$ which maps the exterior of $C_{1-\delta}$ onto $|w|>1$. Thus we have for z on $C_{1-\delta}$

$$|\omega_n(z)| \leqq M_1\Delta^{n+1}(1 - \delta)^{n+1}.$$

The function $\omega_n(z)/w^{n+1}$ is analytic for z in the closed exterior of $C_{1-\delta}$ even at infinity, and w has the modulus unity for z on $C_{1-\delta}$; so we may write for z on and exterior to $C_{1-\delta}$

$$(7.4) \qquad |\omega_n(z)| \leqq M_2\Delta^{n+1}(1 - \delta)^{n+1}|w|^{n+1}.$$

(3) Some of the details of the present proof are due to the referee, replacing incorrect details of our original draft.

We integrate from an arbitrary z_0 on $C_{1-\delta}$ to z on C, choosing as path the image ("Radiusbild") in the z-plane of a radius of the unit circle in the plane of $w=\phi(z)/(1-\delta)$; for all t on C_r, where r is sufficiently near ρ, we have

$$\left|\int_{z_0}^{z}\frac{\omega_n(z)dz}{t-z}\right| \leqq M_3\int_{z_0}^{z}\Delta^{n+1}\left|\phi(z)\right|^{n+1}\left|dz\right| \leqq M_4\int_{z_0}^{z}\Delta^{n+1}\left|\phi(z)\right|^{n+1}\left|\phi'(z)dz\right|$$
$$= M_4\Delta^{n+1}[\left|\phi(z)\right|^{n+2}-\left|\phi(z_0)\right|^{n+2}]/(n+2) \leqq M_5\Delta^{n+2}/(n+2).$$

Here M_5 is independent of n, t, z_0, and z. If a is a fixed point interior to $C_{1-\delta}$, further use of (7.4) yields for z on C

$$\left|\int_{a}^{z}\frac{\omega_n(z)dz}{t-z}\right| \leqq \left|\int_{a}^{z_0}\frac{\omega_n(z)dz}{t-z}\right| + \left|\int_{z_0}^{z}\frac{\omega_n(z)dz}{t-z}\right| \leqq M_6\Delta^{n+2}/(n+2).$$

Also for t on or exterior to C we have $\left|\omega_n(t)\right| \geqq M_7\Delta^{n+1}\left|\phi(t)\right|^{n+1}>0$; hence for z on C we have

$$\left|\int_{a}^{z}[f'(z)-p_n'(z)]dz\right| \leqq M_8\cdot\max\,[\left|f'(t)\right|,\ t \text{ on } C_r]/r^n\cdot n.$$

From Theorem 7.2 it follows that $\left|f'(t)\right| \leqq N(\rho-r)^{p+\alpha}$ *on* C_r; if we put $r=\rho(1-1/n)$ we obtain the conclusion of the theorem.

This reasoning cannot be carried through if the points of interpolation are chosen equidistributed on C itself. Let C be the unit circle $\left|z\right|=1$, whence $\omega_n(z)\equiv z^{n+1}-1$; then for z on C we have

$$\int_{0}^{z}(z^{n+1}-1)dz = \frac{z^{n+2}}{n+2}-z,$$

so no additional factor n appears in the denominator due to the integration. For the particular function $f'(z)\equiv(t-z)^{-1}$, $\left|t\right|=\rho>1$, of class $L(-2,1)$ on C_ρ, with C the circle $\left|z\right|=1$ and $\omega_n(z)=z^{n+1}-1$, the conclusion of Theorem 7.7 is false.

An extension of Theorem 4.2, which likewise extends to higher integrals, is

THEOREM 7.8. *Let C be an analytic Jordan curve. Let $f(z)$ be analytic interior to C_ρ, continuous in the corresponding closed region, and let polynomials $P_n(z)$ exist such that we have on C_ρ*

$$\left|f(z)-P_n(z)\right| \leqq \epsilon_n.$$

Let $p_n(z)$ denote the polynomial which interpolates to $f(z)$ in $n+1$ points equally distributed on $C_{1-\delta}$, $\delta>0$, where $1-\delta$ is sufficiently small. Then we have

$$\left|\int_{a}^{z}[f(z)-p_n(z)]dz\right| \leqq M\epsilon_n/n\rho^n, \qquad z \text{ on } C,$$

where a is an arbitrary point interior to $C_{1-\delta}$, and the integral is taken along an arbitrary path containing no point exterior to C.

The proof here is similar to that of the preceding theorem. Instead of the Lagrange-Hermite interpolation formula we use the following form employed by the authors [1940]

$$f(z) - p_n(z) = \frac{1}{2\pi i}\int_{C_\rho} \frac{\omega_n(z)[f(t) - P_n(t)]}{\omega_n(t)(t-z)}\,dt, \qquad z \text{ on } C, \tag{7.5}$$

where $P_n(z)$ is the polynomial mentioned above and $p_n(z)$ is the polynomial of degree n which interpolates to $f(z)$ in the roots of $\omega_n(z)$, namely, the points z_j on $C_{1-\delta}$ used in the proof of Theorem 7.7. The procedure goes through with only obvious modifications(4).

An analogue of Theorem 4.3 is

THEOREM 7.9. *Let C be an analytic Jordan curve; let $f(z)$ be of class $L(p, \alpha)$, $p \leqq -1$, $0<\alpha\leqq 1$, on C_ρ. Let $p_n(z)$ denote the polynomial of degree n which interpolates to $f(z)$ in $n+1$ points equally distributed on C. Then we have*

$$|f'(z) - p_n'(z)| \leqq M/\rho^n \cdot n^{p+\alpha-1}, \qquad z \text{ on } C.$$

In the formula

$$f(z) - p_n(z) = \frac{1}{2\pi i}\int_{C_r} \frac{\omega_n(z)f(t)}{\omega_n(t)(t-z)}\,dt, \qquad z \text{ on or within } C,\ 1 < r < \rho,$$

let us differentiate with respect to z:

$$f'(z) - p_n'(z) = \frac{1}{2\pi i}\int_{C_r} \frac{f(t)}{\omega_n(t)}\left[\frac{(t-z)\omega_n'(z) + \omega_n(z)}{(t-z)^2}\right] dt.$$

But $|\omega_n(z)| \leqq M\Delta^{n+1}$, z on C, and hence by an extension of Bernstein's theo-

(4) We mention here the following theorem, analogous to Theorem 7.8:

Let C be an analytic Jordan curve, let $f(z)$ be analytic interior to C and continuous in the corresponding closed region, and let there exist polynomials $P_n(z)$ such that we have for z on C_ρ

$$|f(z) - P_n(z)| \leqq \epsilon_n.$$

Then we have

$$|f(z) - a_0p_0(z) - a_1p_1(z) - \cdots - a_np_n(z)| \leqq M\epsilon_n/\rho^n, \qquad z \text{ on } C,$$

where $\sum a_kp_k(z)$ is the expansion of $f(z)$ in Faber polynomials belonging to C.

This theorem follows from the formulas (in the notation of Faber [1920])

$$a_k = \frac{1}{2\pi i}\int_{C_\rho} \frac{f(t)}{w^{k+1}}\,dt = \frac{1}{2\pi i}\int_{C_\rho} \frac{f(t) - P_n(t)}{w^{k+1}}\,dt, \qquad k > n;$$

for the expansion of $P_n(z)$ in Faber polynomials is unique, whence

$$\int_{C_\rho} \frac{P_n(t)}{w^{k+1}}\,dt = 0, \qquad k > n.$$

rem [Sewell, 1937] it follows that $|\omega_n'(z)| \leqq M\Delta^{n+1}(n+1)$, z on C. Thus we have for z on C

$$|f'(z) - p_n'(z)| \leqq M_1 n \cdot \max [|f(t)|, \ t \text{ on } C_r]/r^{n+1};$$

by using Theorem 7.2 and putting $r=\rho(1-1/n)$ we obtain the inequality of the theorem; the proof is complete.

A direct analogue of Theorem 4.4 is

THEOREM 7.10. *Let $f(z)$ be analytic interior to the analytic Jordan curve C, continuous in the corresponding closed region. Let there exist polynomials $P_n(z)$ such that we have on C_ρ*

$$|f(z) - P_n(z)| \leqq \epsilon_n.$$

Let $p_n(z)$ denote the polynomial of degree n which interpolates to $f(z)$ in $n+1$ points equally distributed on C. Then we have

$$|f'(z) - p_n'(z)| \leqq Mn\epsilon_n/\rho^n, \qquad z \text{ on } C.$$

The proof here is similar to that of the preceding theorem except that we use the formula (7.5). The details are left to the reader.

Theorem 7.10 includes the conclusion of Theorem 7.9 with the restriction $p \geqq 0$, $0<\alpha\leqq 1$; in the boundary case $p+\alpha=0$ we set $\epsilon_n = M_0$.

The theorems just established are the analogues of those of §§3 and 4; the latter are limiting cases but not properly special cases of the former. In the converse direction we have the following analogue of Theorem 5.1:

THEOREM 7.11. *Let C be an analytic Jordan curve and let $f(z)$ be defined on C. For each n, $n=1, 2, \cdots$, let a polynomial $p_n(z)$ exist such that*

$$|f(z) - p_n(z)| \leqq M/\rho^n \cdot n^{p+\alpha+1}, \qquad z \text{ on } C,\ \rho > 1.$$

Then $f(z)$, when suitably defined, is analytic interior to C_ρ and is of class $L(p, \alpha)$ on C_ρ if $p+\alpha+1$ is not a positive integer, and of class $L'(p, \alpha)$ if $p+\alpha+1$ is a positive integer.

The extension of Bernstein's lemma [Walsh, 1935, p. 77] applies here and there are no essential changes necessary in the proof of Theorem 5.1 as given, except that in the case where $p+\alpha+1$ is a positive integer we now use the polynomials $q_n(z)$ normal and orthogonal on C. The function $f(z)$ is analytic throughout the interior of C_ρ [Walsh, 1935, p. 78], and we have $f(z) \equiv \sum_{k=0}^{\infty} a_k q_k(z)$ throughout the interior of C_ρ, uniformly on any closed set interior to C_ρ. By virtue of the given $p_n(z)$ and the least-square property of the $q_n(z)$ we have

$$|a_n| \leqq M_1/\rho^n \cdot n^{p+\alpha+1}.$$

The polynomials $q_n(z)$ are uniformly bounded on C [Szegö, 1939, p. 365], so

by Bernstein's inequality in an extended form we have

$$\left| q_n^{(p+\alpha+1)}(z) \right| \leqq M_2 n^{p+\alpha+1}, \qquad z \text{ on } C,$$

$$\left| f^{(p+\alpha+1)}(z) \right| \leqq \sum_{k=0}^{\infty} M_3 \cdot r^k/\rho^k, \qquad z \text{ on } C_r;$$

the reasoning used in connection with (5.4) now applies.

A consequence of such theorems as 7.5–7.10 is inequalities on degree of approximation of polynomials *of best approximation in the sense of Tchebycheff.* For such polynomials and others (e.g., as in Theorem 7.11), Problem γ, namely, the study of degree of convergence on C_σ, $1<\sigma<\rho$, can be treated by the methods that we have already developed.

It is to be observed that the methods we use in §7 apply to much more general measures of degree of convergence and asymptotic conditions than those exhibited by functions of class $L(p, \alpha)$. In fact we have the following theorems:

THEOREM 7.12. *Let C be an analytic Jordan curve and suppose on each C_r for which r lies in an interval $r_0<r<\rho$ we have $|f(z)| \leqq \phi(\rho-r)$, where the function $\phi(x)$ is defined in some interval $0<x<x_0$. Then there exist polynomials $p_n(z)$ such that we have*

$$\left| f(z) - p_n(z) \right| \leqq \frac{M \cdot \phi(\rho/n)}{\rho^n}, \qquad z \text{ on } C.$$

The usual results hold also for approximation to integrals and derivatives of $f(z)$; additional factors n appear on the right in denominator or numerator.

THEOREM 7.13. *Let $f(z)$ be defined on C and polynomials $p_n(z)$ exist such that*

$$\left| f(z) - p_n(z) \right| \leqq \frac{\phi(1/n)}{\rho^n}, \qquad z \text{ on } C,$$

where $\phi(x)$ is defined and monotonic throughout some interval $0<x<x_0 \geqq 1$. Then we have

$$\left| f(z) \right| \leqq M \sum_{n=1}^{\infty} \phi(1/n) r^n/\rho^n, \qquad z \text{ on } C_r,\ r<\rho,$$

provided this series converges.

8. **Approximation on a line segment. Trigonometric approximation.** Approximation on a finite line segment is analogous to approximation on an analytic Jordan curve, provided the approximated function is analytic on the given segment. In §7 we studied approximation on Jordan curves by interpolation in equally distributed points; these same points serve in the study of approximation on a line segment.

The roots of the polynomial $T_n(z)=2^{-n+1}\cos(n\cos^{-1}z)$ are equally distributed on the segment C: $-1\leqq z\leqq 1$. The inequality

$$\left|\frac{T_n(z)}{T_n(t)}\right| \leqq \frac{M}{\rho^n}, \qquad z \text{ on } C,\ t \text{ on } C_\rho,$$

is known [for instance Walsh and Sewell, 1940]. We may choose M independent of ρ, for $\rho>\rho_0>1$. Consequently the discussion of §7 concerning direct approximation on C is valid in the present case; we state

THEOREM 8.1. *If C is the segment $-1\leqq z\leqq 1$ and $f(z)$ is analytic interior to C_ρ and of class $L(p, \alpha)$ or $L'(p, \alpha)$ on C_ρ, then there exist polynomials $p_n(z)$ such that we have on C*

$$|f(z) - p_n(z)| \leqq M/\rho^n \cdot n^{p+\alpha}.$$

For the case $p\geqq 0$, Theorem 8.1 has already been established [Walsh and Sewell, 1940]; for the case $p<0$, the proof follows that of Theorem 3.1. The polynomials $p_n(z)$ are chosen as the polynomials interpolating to $f(z)$ in the zeros of $T_{n+1}(z)$. For the class $L'(p, \alpha)$, compare the remarks on integration below.

In the direction of a converse we have

THEOREM 8.2. *Let C be the segment $-1\leqq z\leqq 1$ and let $f(z)$ be defined on C. For each n, $n=1, 2, \cdots$, let a polynomial $p_n(z)$ exist such that*

$$|f(z) - p_n(z)| \leqq M/\rho^n \cdot n^{p+\alpha+1}, \qquad z \text{ on } C,\ \rho > 1.$$

Then $f(z)$ when suitably defined, is of class $L(p, \alpha)$ on C_ρ if $p+\alpha+1$ is not a positive integer, and of class $L'(p, \alpha)$ if $p+\alpha+1$ is a positive integer.

The proof of Theorem 8.2 is essentially the same as that of Theorem 5.1 for $p+\alpha+1$ not a positive integer; for $p+\alpha+1$ a positive integer we proceed as in Theorem 7.11, using polynomials normal and orthogonal on a particular C_σ, $1<\sigma<\rho$, and the inequality

$$|f(z) - p_n(z)| \leqq M_1\sigma^n/\rho^n n^{p+\alpha+1}, \qquad z \text{ on } C_\sigma.$$

The entire discussion of §7 concerning differentiation and integration of sequences remains essentially valid, except that in differentiation the additional factor n is to be replaced by n^2; on the segment $-1\leqq z\leqq 1$, we have $|T_n'(z)|\leqq n^2/2^{n-1}$.

In the study of integration of sequences (compare Theorems 7.7 and 7.8) we use the following evaluation. From the interpolation formula

$$f(z) - p_{n-1}(z) = \frac{1}{2\pi i}\int_{C_r} \frac{T_n(z)}{T_n(t)}\, \frac{f(t) - P_{n-1}(t)}{t - z}\, dt, \qquad -1 \leqq z \leqq 1,$$

we have by integration

$$\int_0^z [f(z) - p_{n-1}(z)]dz = \frac{1}{2\pi i}\int_{C_r} \frac{[f(t) - P_{n-1}(t)]}{T_n(t)}\left[\int_0^z \frac{T_n(z)}{t-z}dz\right]dt.$$

Thus we have to consider merely

$$\int_0^z \frac{\cos(n\cos^{-1} z)dz}{t-z} = \frac{1}{t-z}\int \cos(n\cos^{-1} z)dz\Big]_0^z$$

$$- \int_0^z \left[\int \cos(n\cos^{-1} z)dz\right]\frac{dz}{(t-z)^2};$$

for all t on C_r and z on C we have

$$\left|\int_0^z \frac{\cos(n\cos^{-1} z)}{t-z}dz\right| \leqq M_1/n.$$

Consequently we obtain the same inequalities for integration of sequences in the case of a line segment C as for C an analytic Jordan curve.

This completes our study of the line segment. It is of interest to note that Theorems 8.1 and 8.2 might have been proved by mapping the complement of C conformally on the exterior of the unit circle γ: $|w|=1$, and applying the results already obtained (Theorems 3.1, 3.2, 5.1) for the unit circle. However the above method is more direct.

We now consider the unit circle and functions analytic in the annulus γ_ρ: $\rho>|z|>1/\rho<1$. Suppose $f(z)=\sum_{m=-\infty}^{m=\infty} c_m z^m$ is analytic in γ_ρ; it is well known that we may write $f(z)\equiv f_1(z)+f_2(z)$, where

$$f_1(z) = \sum_{m=0}^{\infty} c_m z^m, \qquad |z| < \rho, \qquad f_2(z) = \sum_{m=-1}^{-\infty} c_m z^m, \qquad |z| > 1/\rho.$$

If $f_1(z)$ and $f_2(1/z)$ belong to the class $L(p, \alpha)$ or $L'(p, \alpha)$ on $|z|=\rho$ we say that $f(z)$ belongs to the class $L(p, \alpha)$ or $L'(p, \alpha)$ on γ_ρ. With this definition it is easy to establish theorems analogous to Theorems 3.1, 3.2, and 5.1.

THEOREM 8.3. *Let $f(z)$ belong to the class $L(p, \alpha)$ or $L'(p, \alpha)$ in the annular region $\rho>|z|>1/\rho<1$, and let $f(z)=\sum_{m=-\infty}^{\infty} c_m z^m$. Then with the notation $a_m=c_m+c_{-m}$, $b_m=i(c_m-c_{-m})$ we have the relation*

$$\left|f(e^{i\theta}) - \left[\frac{a_0}{2} + \sum_{m=1}^{n}(a_m\cos m\theta + b_m \sin m\theta)\right]\right| \leqq M/\rho^n\cdot n^{p+\alpha}.$$

In the converse direction we are concerned with a polynomial $p_n(z, 1/z)$ of degree n in z and $1/z$, namely a function of the form

$$p_n(z, 1/z) \equiv a_{-n}z^{-n} + \cdots + a_0 + \cdots + a_n z^n.$$

THEOREM 8.4. *Let $f(z)$ be defined on $|z|=1$ and let polynomials $p_n(z, 1/z)$ exist such that*

$$|f(z) - p_n(z, 1/z)| \leqq M/\rho^n \cdot n^{p+\alpha+1}, \qquad |z| = 1, \rho > 1.$$

Then $f(z)$, if properly extended from the unit circle, belongs to the class $L(p, \alpha)$ in the annulus $\rho > |z| > 1/\rho$ if $p+\alpha+1$ is not a positive integer, and to the class $L'(p, \alpha)$ if $p+\alpha+1$ is a positive integer.

For $p \geqq 0$, $\alpha < 1$ Theorems 8.3 and 8.4 have already been proved [Walsh and Sewell, 1938]; for $p<0$, $0<\alpha\leqq 1$, and for $p\geqq 0$, $\alpha=1$ the methods for the unit circle may be applied to $f_1(z)$ and $f_2(z)$; in the latter case we make use of S. Bernstein's theorem concerning the derivative of a trigonometric polynomial of order n. Theorems 8.3 and 8.4 may be interpreted as results on trigonometric approximation (loc. cit.); in fact the transformation $w = e^{iz}$ suggests directly the definitions involved in the following theorems; formal definitions and proofs may be easily supplied by the reader:

THEOREM 8.5. *Let the function $f(z)$ be periodic with period 2π and of class $L(p, \alpha)$ or $L'(p, \alpha)$ in the band $|y| < \log \rho > 0$, $z = x+iy$. Then there exist trigonometric polynomials $t_n(z)$ such that we have for all real z*

$$|f(z) - t_n(z)| \leqq M/\rho^n \cdot n^{p+\alpha}.$$

THEOREM 8.6. *Let the function $f(z)$ be defined for all real z and periodic with period 2π. Let trigonometric polynomials $t_n(z)$ exist such that for all real $z = x+iy$*

$$|f(z) - t_n(z)| \leqq M/\rho^n \cdot n^{p+\alpha+1}.$$

Then $f(z)$ belongs to the class $L(p, \alpha)$ on $|y| < \log \rho$ if $p+\alpha+1$ is not a positive integer, and to the class $L'(p, \alpha)$ if $p+\alpha+1$ is a positive integer.

Results analogous to those of the present section have already been established by de la Vallée Poussin [1919] and S. Bernstein [1926], who study approximation by trigonometric polynomials and approximation on the segment $(-1, 1)$, for the case that the function $f(z)$ has only isolated singularities.

9. **Approximation measured by an integral.** Well known methods apply to our results of §§3–8 on approximation, and give us theorems on approximation by polynomials as measured by line integrals. For instance under the hypothesis of Theorem 7.5 or 8.1 there exist polynomials $p_n(z)$ such that we have

$$\int_C |f(z) - p_n(z)|^m \cdot |dz| \leqq M/\rho^{mn} n^{m(p+\alpha)}, \qquad m > 0. \tag{9.1}$$

Conversely an inequality of form (9.1) implies that $f(z)$ is of class $L(p-1, \alpha)$ on C_ρ if $p+\alpha$ is not a positive integer and of class $L'(p-1, \alpha)$ on C_ρ if $p+\alpha$ is a positive integer; but of course when (9.1) is given, the function $f(z)$ appears in our hypothesis merely almost everywhere, and the characterization of $f(z)$ just given contemplates a revision of the definition of $f(z)$ on a set of measure zero.

The statements just made have already been established [Walsh and Sewell, 1940, 1940a] for the case $p+\alpha-1>0$, $p+\alpha$ not a positive integer, and can be established for the remaining case by standard methods [Walsh, 1935, p. 92]; compare the proof of Theorem 8.2.

These remarks on approximation as measured by an integral apply likewise if a suitably restricted norm function is introduced.

10. **Integrated Lipschitz conditions and integral asymptotic conditions.** We described in §2 a classification of functions based on results of Hardy and Littlewood, a classification which we have seen (§§3–5) to be highly appropriate in the study of both direct and indirect theorems under Problem β. Still another classification, likewise based on results of Hardy and Littlewood, is of interest and also appropriate in the study of Problem β. But this new classification is far less elementary and intuitive than the former one, and also has been far less used; for this reason we have emphasized the one rather than the other. Nevertheless the more sophisticated classification deserves some treatment, which we proceed to develop in the special case of the circle, and to apply in the study of approximation.

If the function $f(z)$ is analytic for $|z|<1$, we use the definition ($|z|=r<1$)

$$M_m = M_m(f) = M_m(r, f) = \left(\frac{1}{2\pi}\int_{-\pi}^{\pi} |f(re^{i\theta})|^m d\theta\right)^{1/m};$$

this has a meaning for every $m>0$, but is to be used below primarily for $m=2$.

We shall say that the function $f(z)$ analytic for $|z|<1$ is *of class* $L_2(p, \alpha)$, where p is a negative integer and $0<\alpha\leqq 1$, provided we have

$$M_2(f) \leqq M(1-r)^{p+\alpha}. \tag{10.1}$$

We shall say that the function $f(z)$ analytic for $|z|<1$ and with boundary values almost everywhere on $|z|=1$ is *of class* $L_2(0, \alpha)$, $0<\alpha\leqq 1$ provided there is satisfied the integrated Lipschitz condition of order α:

$$\frac{1}{2\pi}\int_{-\pi}^{\pi} |f(e^{i\theta+ih}) - f(e^{i\theta})|^2 d\theta \leqq Mh^{2\alpha}. \tag{10.2}$$

With these definitions, Hardy and Littlewood [1932] prove three important theorems:

THEOREM 10.1. *If $p+\alpha\leqq 0$ and if $f(z)$ is of class $L_2(p, \alpha)$, then $f^{(k)}(z)$ is of class $L_2(p-k, \alpha)$.*

THEOREM 10.2. *If $p+\alpha<0$, $p-k+\alpha<0$, and if $f(z)$ is of class $L_2(p-k, \alpha)$, then the kth integral of $f(z)$ is of class $L_2(p, \alpha)$.*

THEOREM 10.3. *A necessary and sufficient condition that $f(z)$ be of class $L_2(0, \alpha)$, $0<\alpha\leqq 1$, is that $f'(z)$ be of class $L_2(-1, \alpha)$.*

It is now natural to say that $f(z)$ is of class $L_2(p, \alpha)$, where p is a non-negative integer and $0<\alpha\leqq 1$ provided $f^{(p+1)}(z)$ is of class $L_2(-1, \alpha)$. With this understanding we have at once for every p and k

THEOREM 10.4. *If $f(z)$ is of class $L_2(p, \alpha)$, then the function $f^{(k)}(z)$ is of class $L_2(p-k, \alpha)$ and if $(p+\alpha)(p+k+\alpha+1)$ is not a negative integer the kth integral of $f(z)$ is of class $L_2(p+k, \alpha)$.*

As in §2, there is here an exception if $\alpha=1$. We define the class $L_2'(-1, 1)$ as the class of integrals of functions of class $L_2(-2, 1)$, and the class $L_2'(p, 1)$ as the class of $(p+2)$th iterated integrals of functions of the class $L_2(-2, 1)$, where $p>-2$. It follows that if $f(z)$ is of class $L_2'(p, 1)$, with $p>-2$, then $f^{(k)}(z)$, $0<k<p+2$, is of class $L_2'(p-k, 1)$; also $f^{(k)}(z)$, $k\geqq p+2$, is of class $L_2(p-k, 1)$; if $f(z)$ is of class $L_2(p, 1)$, $p<-2$, then $f^{(k)}(z)$ is of class $L_2(p-k, 1)$ and the kth iterated integral of $f(z)$ is of class $L_2(p+k, 1)$ or $L_2'(p+k, 1)$ according as $p+k\leqq -2$ or $p+k>-2$.

These preliminaries completed we are in position to study approximation:

THEOREM 10.5. *If $f(z)$ is of class $L_2(p, \alpha)$ or of class $L_2'(p, \alpha)$, there exist polynomials $p_n(z)$ such that we have for $|z|=1/\rho<1$*

$$|f(z) - p_n(z)| \leqq M/\rho^n \cdot n^{p+\alpha}. \tag{10.3}$$

In the case $p+\alpha\leqq 0$, the method previously given (§3) is applicable; we employ (10.1) and the Schwarz inequality; in the case $p+\alpha>0$, we use that same method but applied now to the function $f^{(p+1)}(z)$, and integrate $p+1$ times under the integral sign in the interpolation formula.

We present an alternative proof of Theorem 10.5 for the class $L_2(p, \alpha)$, $p\geqq 0$. If $f(z)=\sum a_n z^n$ is of class $L_2(0, \alpha)$ we may write [this method is well known]

$$\begin{aligned} a_n &= \frac{1}{2\pi i}\int_{|z|=1} \frac{f(z)}{z^{n+1}}\, dz = \frac{1}{2\pi}\int_0^{2\pi} \frac{f(e^{i\theta})}{e^{in\theta}}\, d\theta \\ &= \frac{-1}{2\pi}\int_0^{2\pi} \frac{f(e^{i(\theta+\pi/n)})}{e^{in\theta}}\, d\theta = \frac{1}{4\pi}\int_0^{2\pi} [f(e^{i\theta}) - f(e^{i(\theta+\pi/n)})]\frac{d\theta}{e^{in\theta}}, \end{aligned}$$

whence by Schwarz's inequality and the fundamental definition of class $L_2(0, \alpha)$, we have $|a_n|\leqq M_0/n^\alpha$. If $f(z)=\sum a_n z^n$ is of class $L_2(p, \alpha)$, $p>0$, we have by p-fold differentiation and use of the preceding relation, $|a_n| \leqq M'/n^{p+\alpha}$(5). Consequently on the circle $|z|=1/\rho<1$ we have

$$\left| f(z) - \sum_{\nu=0}^{n} a_\nu z^\nu \right| \leqq \sum_{n+1}^{\infty} |a_\nu z^\nu| \leqq M'\sum_{n+1}^{\infty} \frac{1}{\rho^\nu \nu^{p+\alpha}} \leqq \frac{M'}{n^{p+\alpha}}\sum_{n+1}^{\infty}\frac{1}{\rho^\nu} \leqq \frac{M_1}{n^{p+\alpha}\rho^n},$$

which establishes (10.3) for the case $p\geqq 0$.

(5) This last inequality is readily proved for functions of class $L(p, \alpha)$, $L'(p, \alpha)$, $L_2(p, \alpha)$, and $L_2'(p, \alpha)$, for every p and $0<\alpha\leqq 1$.

The indirect approximation problem is similarly handled:

THEOREM 10.6. *Let there exist polynomials $p_n(z)$ such that*

$$|f(z) - p_n(z)| \leq M/\rho^n \cdot n^{p+\alpha+1/2} \tag{10.4}$$

is valid for $|z| = 1/\rho < 1$; then $f(z)$ is of class $L_2(p, \alpha)$ if $p+\alpha+1$ is not a positive integer and is of class $L_2'(p, \alpha)$ if $p+\alpha+1$ is a positive integer.

Our proof of Theorem 10.6 uses not (10.4) directly, but the inequality

$$\sum_{\nu=n+1}^{\infty} |a_\nu|^2/\rho^{2\nu} \leq M_1/\rho^{2n}n^{2p+2\alpha+1}, \qquad f(z) = \sum_{\nu=0}^{\infty} a_\nu z^\nu, \tag{10.5}$$

which is a direct consequence of (10.4) by virtue of the least-square property of the polynomials $s_n(z) = \sum_{\nu=0}^{n} a_\nu z^\nu$ on the circle $|z| = 1/\rho$:

$$\sum_{\nu=n+1}^{\infty} |a_\nu|^2/\rho^{2\nu} = \frac{\rho}{2\pi}\int_{|z|=1/\rho} |f(z) - s_n(z)|^2 |dz|$$

$$\leq \frac{\rho}{2\pi}\int_{|z|=1/\rho} |f(z) - p_n(z)|^2 |dz|.$$

An inequality which follows from (10.5) is

$$|a_n|^2 \leq M_2/n^{2p+2\alpha+1}. \tag{10.6}$$

Let us now choose the non-negative integer k in such a manner that we have $2k > 2p+2\alpha+1$; we have from (10.6)

$$\begin{gathered} f^{(k)}(z) = \sum_{\nu=0}^{\infty} b_\nu z^\nu, \qquad b_\nu = (\nu + k)(\nu + k - 1) \cdots (\nu + 1)a_{\nu+k}, \\ |b_n|^2 \leq M_3 n^{2k-2p-2\alpha-1}. \end{gathered} \tag{10.7}$$

A consequence of (10.7) is (see below)

$$\begin{aligned} \frac{1}{2\pi r}\int_{|z|=r} |f^{(k)}(z)|^2 |dz| &= \sum_{n=0}^{\infty} |b_n|^2 r^{2n} \leq M_3 \sum_{n=0}^{\infty} n^{2k-2p-2\alpha-1} r^{2n} \\ &\leq M_4(1 - r^2)^{2(p+\alpha-k)}, \end{aligned} \tag{10.8}$$

whence $f^{(k)}(z)$ is of class $L_2(p-k, \alpha)$. By Theorem 10.4 the kth integral of $f^{(k)}(z)$ is of class $L_2(p, \alpha)$ if $p+\alpha+1$ is not a positive integer; and the kth integral of $f^{(k)}(z)$ is of class $L_2'(p, \alpha)$ if $p+\alpha+1$ is a positive integer, so Theorem 10.6 is established.

It remains to justify the last inequality in (10.8); this is accomplished by the method used in the treatment of inequality (5.4).

It has been noted that our proof of Theorem 10.6 uses not (10.4) as hy-

pothesis, but rather (10.5). It is of interest to remark that *the hypothesis may be taken as*

$$\int_{|z|=1/\rho} |f(z) - p_n(z)|^m |dz| \leqq M/\rho^{mn} n^{m(p+\alpha+1/2)}, \tag{10.9}$$

where m is an arbitrary positive number. For (10.9) implies by standard algebraic inequalities (e.g., Walsh [1935, p. 93])

$$\int_{|z|=1/\rho} |p_{n+1}(z) - p_n(z)|^m |dz| \leqq M_1/\rho^{mn} n^{m(p+\alpha+1/2)},$$

which in turn yields (Walsh [1935, p. 92])

$$|p_{n+1}(z) - p_n(z)| \leqq M_2/\rho_1^n n^{p+\alpha+1/2} \tag{10.10}$$

on the circle $|z| = 1/\rho_1$, with $1 < \rho_1 < \rho$. Inequality (10.10) implies by the method of proof of (10.5)

$$\sum_{\nu=n+1}^{\infty} |a_\nu|^2/\rho_1^{2\nu} \leqq M_3/\rho_1^{2n} n^{2p+2\alpha+1},$$

which is precisely of form (10.5) with ρ replaced by ρ_1, and which suffices to prove the conclusion of Theorem 10.6.

We mention the following beautiful result, stated without proof by Hardy and Littlewood [1928]: *The class of functions $f(\theta)$ satisfying an integrated Lipschitz condition of order α is identical with the class of functions which can be approximated in the mean by trigonometric polynomials of degree n with error not greater than M/n^α.* It is to be noted that Theorems 10.5 and 10.6 have been proved without the help of this result. Nevertheless this result can readily be used to give a new proof of Theorem 10.5 for the case $p \geqq 0$, by the methods already developed by the authors [1940].

Theorems 10.5 and 10.6 are obviously to be compared with Theorems 3.1 and 5.1. The discrepancy between the exponents of n in (10.3) and (10.4) is only $\frac{1}{2}$, whereas that between the exponents of n in (3.1) and (5.1) is unity, so in this respect Theorems 10.5 and 10.6 are an improvement over Theorems 3.1 and 5.1. It may be remarked, however, that the proof of Theorem 10.6 as given does not admit of direct extension to an arbitrary analytic Jordan curve C.

Theorems 10.5 and 10.6 are in a sense the best possible results, namely in the sense that we cannot replace $p+\alpha$ in (10.3) by any $\alpha' > p+\alpha$, and that we cannot replace $p+\alpha$ in (10.4) by any $\alpha'' < p+\alpha$; we proceed to illustrate this fact by specific examples.

For an example in connection with Theorem 10.5 we set

$$f(z) = \sum_{n=1}^{\infty} 2^{\beta n} z^{2^n}, \qquad \beta > 0;$$

we have for $|z| = r$

$$\left|\sum_{n=1}^{\infty} 2^{\beta n} z^{2^n}\right| \leqq \int_1^{\infty} (2^x)^{\beta} r^{2^x} dx \leqq M(1-r)^{-\beta},$$

and hence a fortiori $f(z)$ is of class $L_2(p, \alpha)$ where $0<\alpha\leqq 1$, p is a negative integer, and $-\beta=p+\alpha$. Also it follows by the method used in §5 that there exist no polynomials $p_n(z)$ such that for every n we have

$$|f(z) - p_n(z)| \leqq M/\rho^n n^{p+\alpha+\delta}, \qquad \delta > 0, \ |z| = 1/\rho < 1.$$

Thus we see that in the sense mentioned, Theorem 10.5 cannot be improved for $p+\alpha\leqq 0$; by integration and Theorems 10.2 and 10.3 the conclusion extends to non-integral positive $p+\alpha$. The case of integral non-negative $p+\alpha$ can be treated as in §5.

For an example in connection with Theorem 10.6 we choose

$$f(z) = (1-z)^{\beta-1/2} = \sum_{m=0}^{\infty} a_m z^m, \qquad \beta \leqq 0,$$

where [e.g., de la Vallée Poussin, 1914, §399]

$$\frac{M_2}{m^{\beta+1/2}} \leqq |a_m| \leqq \frac{M_1}{m^{\beta+1/2}}, \qquad M_2 > 0. \tag{10.11}$$

In §5 we have seen that

$$\left|f(z) - \sum_{m=0}^{n} a_m z^m\right| \leqq M_3/n^{\beta+1/2}\rho^n, \qquad |z| = 1/\rho.$$

But we have

$$\frac{1}{2\pi r}\int_{-\pi}^{\pi} |1-z|^{2\beta-1} d\theta = \sum_{m=0}^{\infty} |a_m|^2 r^{2m}, \qquad |z| = r,$$

and hence by inequality (10.11) we see that $f(z)$ is of class $L_2(p, \alpha)$, if $\beta=p+\alpha<0$, and of class $L_2'(p, \alpha)$ if $\beta=p+\alpha=0$; in each case $f(z)$ is of no higher class. The same method as above serves to extend the scope of the example to all values of $p+\alpha$. Consequently Theorems 10.5 and 10.6 are the best possible in the sense mentioned.

It may be observed that for $\delta>1$ the Hölder inequality

$$\int |f| \leqq M\left(\int |f|^{\delta}\right)^{1/\delta}$$

and for $\delta=1$ more elementary inequalities establish the conclusion of Theorem 10.5, where now $f(z)$ is an arbitrary function of class $L_\delta(p, \alpha)$ or $L_\delta'(p, \alpha)$;

suitable definitions of these classes are fairly obvious from our previous definitions. We remark that $L_s(-1, 1)$ is identical with the class of functions H_s studied by F. Riesz [1923], namely functions $f(z)$ analytic for $|z|<1$ such that

$$\int_{|z|=r} |f(z)|^s |dz|$$

is uniformly bounded for all $r<1$.

Theorems 10.5 and 10.6 extend at once to the situations of §8. Theorem 10.5 extends also to the case that C is an arbitrary analytic Jordan curve; but the writers have not as yet extended Theorem 10.6 to this more general case. We have in the present paper insisted on ordinary Lipschitz conditions and asymptotic conditions rather than on integrated Lipschitz conditions and mean asymptotic conditions because the theory of the latter concepts is not as yet widely developed, and because the former concepts are relatively simple and more direct.

11. **Direct methods on Problems α and β.** In the above sections we have established various results on Problem β, results which are as favorable in many respects as can be obtained. On the other hand, our methods have been in part relatively high-powered, for instance in our proofs of Theorems 7.7 and 7.8. However, some results only slightly less favorable than those obtained above and elsewhere can be established by thoroughly immediate and elementary methods, with a minimum of machinery, as we now proceed to indicate for both Problem α(6) and Problem β. For the present we restrict ourselves to the case of functions analytic in the unit circle γ: $|z|=1$.

If $f(z)$ is of class $L(p, \alpha)$, $p+\alpha \leqq 0$, our results as already established (Theorem 3.1) for approximation on $|z|=1$ are obtained by elementary methods; these results refer to Problem β, and Problem α does not properly present itself.

If $f(z)$ is assumed to satisfy a uniform *radial* Lipschitz condition we can also proceed by elementary methods:

THEOREM 11.1. *Let $f(z)$ be analytic and bounded for $|z|<1$ and satisfy a uniform radial Lipschitz condition, in the sense*

$$|f(e^{i\theta}) - f(re^{i\theta})| \leqq M(1-r)^\alpha, \qquad 0<r<1, \tag{11.1}$$

where M is independent of r and θ. Then there exist polynomials $p_n(z)$ such that we have for all θ

$$|f(e^{i\theta}) - p_n(e^{i\theta})| \leqq M'(\log n)^\alpha/n^\alpha; \tag{11.2}$$

(6) For the set C: $|z|\leqq 1$, Problem α is the study of degree of approximation on C of a function analytic in $|z|<1$ satisfying given conditions of continuity on or in the neighborhood of γ: $|z|=1$.

indeed, the $p_n(z)$ *may be defined as* $s_n(r_nz)$, *where* $s_n(z)$ *is the sum of the first* $n+1$ *terms of the Taylor development of* $f(z)$ *and* $r_n = 1-(2\log n)/n$.

We set $f(z)=\sum_{k=0}^{\infty}a_kz^k$, $s_n(z)=\sum_{k=0}^{\infty}a_kz^k$, and by the boundedness of $f(z)$

$$|f(re^{i\theta}) - s_n(re^{i\theta})| = \left|\frac{1}{2\pi i}\int_{|t|=1}\frac{(re^{i\theta})^{n+1}}{t^{n+1}}\frac{f(t)}{(t-z)}dt\right| \leqq \frac{M_1r^{n+2}}{1-r}, \qquad r<1,$$

for $z=re^{i\theta}$, where M_1 is a constant depending only on $f(z)$. We have from (11.1) by addition

$$|f(e^{i\theta}) - s_n(re^{i\theta})| \leqq \frac{M_1r^{n+2}}{1-r} + M(1-r)^{\alpha}.$$

Corresponding to each n we choose

$$r = r_n = 1 - 2\log n/n,$$

whence by writing

$$\left(1-\frac{2\log n}{n}\right)^n = \left[\left(1-\frac{2\log n}{n}\right)^{n/2\log n}\right]^{2\log n},$$

which is asymptotic to n^{-2}, we obtain the inequality (11.2).

Theorem 11.1 is a result on Problem α; the corresponding result on Problem β is

THEOREM 11.2. *Let* $f(z)$ *satisfy the hypothesis of Theorem* 11.1. *Then there exist polynomials* $p_n(z)$ *such that we have*

$$|f(z) - p_n(z)| \leqq M''(\log n)^{\alpha}/\rho^n\cdot n^{\alpha}, \qquad |z| = 1/\rho < 1.$$

We write

$$f(z) - p_n(z) = \frac{1}{2\pi i}\int_{|t|=1}\frac{z^{n+1}f(t)}{t^{n+1}(t-z)}dt$$

$$= \frac{1}{2\pi i}\int_{|t|=1}\frac{z^{n+1}f(rt)}{t^{n+1}(t-z)}dt + \frac{1}{2\pi i}\int_{|t|=1}\frac{z^{n+1}[f(t)-f(rt)]}{t^{n+1}(t-z)}dt,$$

$$0<r<1.$$

For each n we choose $r=r_n=1-\log n/n$, whence by the method used in Theorem 11.1 we have the inequality of the theorem.

The method of Theorem 11.2 extends in an elementary way to yield results on approximation to integrals and derivatives of $f(z)$; compare §4.

BIBLIOGRAPHY

BERNSTEIN, S.

1926. *Leçons sur les Propriétés Extrémales et la Meilleure Approximation des Fonctions Analytiques d'une Variable Réelle*, Paris.

CURTISS, JOHN.
1935. *Interpolation in regularly distributed points*, these Transactions, vol. 38, pp. 458–473.
1936. *A note on the degree of polynomial approximation*, Bulletin of the American Mathematical Society, vol. 42, pp. 873–878.

FABER, G.
1920. *Über Tchebyscheffsche Polynome*, Journal für die reine und angewandte Mathematik, vol. 150, pp. 79–106.

HARDY, G. H., AND LITTLEWOOD, J. E.
1928. *A convergence criterion for Fourier series*, Mathematische Zeitschrift, vol. 28, pp. 612-634.
1932. *Some properties of fractional integrals*, ibid., vol. 34, pp. 403–439.
1935. *An inequality*, ibid., vol. 40, pp. 1–40.

RIESZ, F.
1923. *Über die Randwerte einer analytischen Funktion*, ibid., vol. 18, pp. 87–95.

SEWELL, W. E.
1937. *Generalized derivatives and approximation by polynomials*, these Transactions, vol. 41, pp. 84–123.

SZEGÖ, G.
1939. *Orthogonal Polynomials*, American Mathematical Society Colloquium Publications, vol. 23.

DE LA VALLÉE POUSSIN, CH. J.
1914. *Cours d'Analyse*, vol. 1, Paris.
1919. *Leçons sur l'Approximation des Fonctions d'une Variable Réelle*, Paris.

WALSH, J. L.
1935. *Interpolation and Approximation by Rational Functions in the Complex Domain*, American Mathematical Society Colloquium Publications, vol. 20.

WALSH, J. L., AND SEWELL, W. E.
1937. *Note on the relation between continuity and degree of polynomial approximation in the complex domain*, Bulletin of the American Mathematical Society, vol. 43, pp. 557–563.
1938. *Note on degree of trigonometric and polynomial approximation to an analytic function*, ibid., vol. 44, pp. 865–873.
1940. *Sufficient conditions for various degrees of approximation by polynomials*, Duke Mathematical Journal, vol. 6, pp. 658–705.
1940a. *Note on degree of trigonometric and polynomial approximation to an analytic function, in the sense of least pth powers*, Bulletin of the American Mathematical Society, vol. 46, pp. 312–319.

HARVARD UNIVERSITY,
CAMBRIDGE, MASS.,
GEORGIA SCHOOL OF TECHNOLOGY,
ATLANTA, GA.

A THEORY OF ABSOLUTELY CONTINUOUS TRANSFORMATIONS IN THE PLANE

BY

T. RADÓ AND P. REICHELDERFER

CHAPTER I. GENERAL OUTLINE OF THE THEORY

1.1. The concept of an absolutely continuous function of a single real variable plays a fundamental rôle in a variety of one-dimensional problems in geometry and analysis. Many efforts have been made to develop an equally useful concept of *two-dimensional absolute continuity*. The purpose of the present paper is to make a contribution to this field through a comprehensive study of a new type of absolute continuity for continuous transformations in the plane(1). In view of the number and variety of concepts and results, it seems advisable to present first a detailed outline of the theory, in order that the reader may more easily obtain a general idea of it. Thus this chapter contains statements of the fundamental definitions and of the more important theorems, as well as some remarks concerning the relationship of this theory to previous literature. In order to avoid duplications, the rest of the paper consists essentially of concise proofs of the results stated in this chapter.

1.2. We commence with a few remarks concerning the very extensive relevant literature(2) (see Banach [1, 2], Bray [1], Caccioppoli [2], McShane [1], Morrey [1], Rademacher [1], Radó [5], Schauder [1], Young [1, 2, 3]). For definiteness, let us consider a continuous transformation T of the form

$$T: \quad x = x(u, v), \qquad y = y(u, v),$$

where $x(u, v)$, $y(u, v)$ are defined and continuous in the unit square

$$S: \quad 0 \leqq u \leqq 1, \qquad 0 \leqq v \leqq 1.$$

Two essentially different types of two-dimensional absolute continuity have thus far been considered in the literature in the study of such a transformation. The first type, due to Banach (see Banach [2]), is a direct generalization of the concept of an absolutely continuous function $x = x(u)$ of a single real variable (where the relation $x = x(u)$ is interpreted as defining a continuous one-dimensional transformation). If E is any set in the square S and if $T(E)$

Presented to the Society, April 15, 1939, under the title *Some properties of continuous transformations in the plane*; received by the editors June 6, 1940.

(1) The class of transformations which are absolutely continuous, in our sense, in a domain $\mathfrak{D}$ will be denoted by $K_1(\mathfrak{D})$, and will be described in 1.32. Important subclasses of $K_1(\mathfrak{D})$ will be defined in §1.34 and §1.37.

(2) Numbers in square brackets are used to refer to the references listed in the bibliography at the end of this paper.

denotes the image of E in the xy plane, then define[3]

$$\Psi(E) = |T(E)|.$$

If there exists a finite constant M such that

$$\sum \Psi(s_i) < M$$

for every finite sequence of nonoverlapping, closed, oriented[4] squares s_i in S, then T is of bounded variation in the sense of Banach. If, for every $\epsilon > 0$ there exists an $\eta = \eta(\epsilon) > 0$ such that

$$\sum \Psi(s_i) < \epsilon$$

for every finite sequence of nonoverlapping, closed, oriented squares s_i in S satisfying

$$\sum |s_i| < \eta,$$

then T is absolutely continuous in the sense of Banach. For transformations T absolutely continuous in this sense, Banach (see Banach [2]) and Schauder (see Schauder [1]) developed a comprehensive and very beautiful theory which has since been completed in many respects by Radó (see Radó [5]). While this theory is wholly satisfactory from the aesthetic point of view, it is a fact that the underlying concept of absolute continuity is too restrictive to permit extensive applications. Indeed, many important classes of continuous transformations, which have been studied by various mathematicians (cf. §1.3) in recent years, do not possess the type of absolute continuity required by Banach.

1.3. In a majority of recent investigations in this field, the assumptions concerning T read as follows.

(i) The functions $x(u, v)$, $y(u, v)$ are both absolutely continuous in the sense of Tonelli[5].

(ii) Further assumptions, concerning $x(u, v)$, $y(u, v)$ jointly, which we do not have to state. We observe only that these assumptions vary considerably, both in character and in generality.

In contradistinction to the type of absolute continuity described in 1.2, the above condition (i) requires, so to speak, ***iterated linear absolute continuity***

(3) If e be any set, then $|e|$ will denote the exterior measure of e.

(4) A rectangle R, or a square S, is termed oriented when its sides are parallel to the respective coordinate axes. Only such squares and rectangles are considered in this paper.

(5) The function $x(u, v)$, for example, is absolutely continuous in the sense of Tonelli if (i) $x(u, v)$ is of bounded variation in the sense of Tonelli, and (ii) for almost every $v = \eta$, $0 \leqq \eta \leqq 1$, the function $x(u, \eta)$ is absolutely continuous in u, and, for almost every $x = \xi$, $0 \leqq \xi \leqq 1$, the function $x(\xi, v)$ is absolutely continuous in v. Denote by $V(\eta)$ the total variation of $x(u, \eta)$ as a function of u on $0 \leqq u \leqq 1$; denote by $V(\xi)$ the total variation of $x(\xi, v)$ as a function of v on $0 \leqq v \leqq 1$. Then $x(u, v)$ is of bounded variation in the sense of Tonelli if $V(\eta)$ and $V(\xi)$ are summable functions of η and ξ respectively on the interval (0, 1).

of $x(u, v)$ and $y(u, v)$ separately. While results of great importance and generality have been derived from assumptions of this character, it is also a fact that for several of the most advanced results, it has been shown that the assumption (i) above either can be replaced by bounded variation in the sense of Tonelli(⁶), or can be discarded entirely (see Radó [3, 4]).

1.4. The preceding remarks suggest that the concepts of absolute continuity described in §1.2 and §1.3 may not represent the ultimate in usefulness. *The purpose of the present paper is to develop, for continuous transformations in the plane, a theory of absolute continuity which is, in the first place, as truly two-dimensional as the theory initiated by Banach, and which is, in the second place, sufficiently comprehensive to permit us to account for and to improve upon the major results obtained for classes of transformations of the character described in* §1.3.

1.5. The basic idea in our theory is to replace, in the definition of absolute continuity due to Banach (cf. §1.2), the function of squares $\Psi(s) = |T(s)|$ by a (generally) smaller function, making thereby the requirement of absolute continuity less restrictive, and hence the class of absolutely continuous transformations more comprehensive. *Intuitively speaking, we shall replace the measure of the image of the square s by the measure of the essential part of that image.* This idea was derived by Radó from a study of the profound and involved work of Geöcze on the area of continuous surfaces (see Radó [1] for references). After the publication in 1928 of the first applications of this idea (see Radó [1]), essentially the same idea was used in 1930 by Caccioppoli (see Caccioppoli [2]) to develop a theory of absolutely continuous transformations in the plane; he applied his theory, in a series of subsequent papers, to a variety of fundamental problems involving double integrals. Since the work of Caccioppoli is based essentially upon the same original ideas and is pointed toward the same general objectives as our work, the following general remarks seem important.

1.6. The theory with which we are concerned is, fundamentally, *a metric theory of topologically defined set functions.* Such set functions are, as a rule, directly defined only for certain simple sets. A basic feature of the work of Caccioppoli is the additive extension of the range of definition of these set functions to the class of all Borel sets (at least). He infers the possibility of such extensions from one of his theorems which may be described as follows (see Caccioppoli [1]).

Let there be given a function Φ defined for all open sets comprised, say, in a fixed rectangle R and satisfying the following conditions: (i) $\Phi(O) \geqq 0$ for every $O \subset R$; (ii) if $O_1, O_2, \cdots$ are nonoverlapping open sets in R, then $\Phi(\sum O_n) = \sum \Phi(O_n)$; (iii) if $O' \subset O'' \subset R$, then $\Phi(O') \leqq \Phi(O'')$; (iv) if O_ϵ denotes the (open) set of those points in O (where O is any open set in R) whose dis-

(⁶) See (⁵).

tance from the boundary of O exceeds ϵ, then $\lim \Phi(O_\epsilon)=\Phi(O)$ as ϵ tends to zero, for every choice of $O\subset R$. According to Caccioppoli, it should then always be possible to extend the definition of Φ to all Borel sets in R in such a way that the extended function is completely additive on the class of all Borel sets in R.

1.7. Unfortunately, the preceding theorem is false, as the reader will easily verify by considering the following example: let S denote a fixed closed square in the interior of R; define $\Phi(O)=1$ if $O\supset S$, and $\Phi(O)=0$ otherwise. Not only is this general extension theorem of Caccioppoli false, but it is also a fact that the particular set functions which he proposes to extend additively cannot generally be so extended. Since many of his basic definitions are stated in terms of (generally non-existing) additive extensions, we were unable, notwithstanding considerable effort, to understand the details of his theory. However, it would seem that the fundamental error pointed out above prevented Caccioppoli from recognizing essential difficulties in the situation—difficulties which delayed the completion of our own work for several years([7]).

1.8. We now proceed to give a summary of the contents of this paper. Let S denote a bounded point set in the uv-plane. A pair of (real-valued) functions $x(u, v)$, $y(u, v)$, bounded and continuous on S, determines a continuous transformation

$$T: \quad x = x(u, v), \qquad y = y(u, v), \qquad\qquad (u, v) \in S.$$

The symbol $\in$ will be used to mean "is an element of the set." For conciseness, we shall also use the complex notation $z=x+iy$, $w=u+iv$, and write

$$T: \quad z = t(w), \qquad\qquad w \in S,$$

where $t(w)$ is then bounded and continuous on S. We term T a bounded continuous transformation defined on S. If E is any set in the w-plane, then $T(E)$ will denote the image of the set $E\cdot S$ under T in the z-plane. If $\overline{E}$ is any set in the z-plane, then $T^{-1}(\overline{E})$ will denote the set of all those points $w\in S$ whose image is in $\overline{E}$; this set is termed the inverse of the set $\overline{E}$ under T.

1.9. If we are given a second bounded continuous transformation

$$T_*: \quad z = t_*(w), \qquad\qquad w \in S,$$

and if E is any set in the w-plane, then we define

$$\rho(T_*, T; E) = \begin{cases} \text{l.u.b. } |t_*(w) - t(w)| \text{ for } w \in E\cdot S, \text{ if } E\cdot S \neq 0, \\ 0 \text{ if } E\cdot S = 0. \end{cases}$$

This quantity $\rho(T_*, T; E)$ will be called the distance of T_* and T on the set E.

([7]) A detailed analysis of fundamental portions of Caccioppoli's work is contained in the (as yet unpublished) Ohio State University Dissertation, *Some properties of continuous transformations in the plane*, 1939, by Paul V. Reichelderfer.

Clearly this distance is symmetric, satisfies the triangle inequality, and vanishes if and only if $t_*(w) = t(w)$ on $E\cdot S$ (provided $E\cdot S$ is not empty).

1.10. We shall use the symbol $N(z, T, E)$ to denote the number (possibly equal to $+\infty$) of distinct points w in the set $E\cdot T^{-1}(z)$. The set of those points z for which $N(z, T, E) = +\infty$ will be denoted by $\overline{\mathfrak{S}}(\infty, T, E)$.

1.11. The symbol $\mathfrak{R}$ will be used to denote generically a (bounded, finitely connected) Jordan region(8). Given a bounded continuous transformation T defined on a Jordan region $\mathfrak{R}$ in the w-plane (cf. §1.8), and a non-negative integer k, we define the set $\overline{\mathfrak{K}}(k, T, \mathfrak{R})$—termed *the kernel*(9) *of order k of the image of $\mathfrak{R}$ under T* (see Radó [5, p. 202])—as the set of those points z for each of which there exists a number $\epsilon = \epsilon(k, z, T, \mathfrak{R}) > 0$ such that, for every continuous transformation T_* defined on $\mathfrak{R}$ and satisfying the inequality $\rho(T_*, T; \mathfrak{R}) < \epsilon$ (cf. §1.9), it is true that $N(z, T_*, \mathfrak{R}) \geqq k$ (cf. §1.10). Clearly a kernel of order $k > 0$ is a subset of every kernel of lower order. Next, we set(10)

$$\overline{\mathfrak{K}}(\infty, T, \mathfrak{R}) = \prod_{k=0}^{\infty} \overline{\mathfrak{K}}(k, T, \mathfrak{R}).$$

This set will be called the kernel of order $+\infty$. Some, or all of the kernels of order $k \geqq 1$ may be empty.

1.12. In the z-plane, we now define a function $\kappa(z, T, \mathfrak{R})$—called *the essential multiplicity of the point z in the image of $\mathfrak{R}$ under T*—as follows (see Radó [5, p. 205]):

$$\kappa(z, T, \mathfrak{R}) = \begin{cases} k & \text{for } z \in \overline{\mathfrak{K}}(k, T, \mathfrak{R}) - \overline{\mathfrak{K}}(k+1, T, \mathfrak{R}), \\ +\infty & \text{for } z \in \overline{\mathfrak{K}}(\infty, T, \mathfrak{R}). \end{cases}$$

Obviously $\kappa(z, T, \mathfrak{R}) \leqq N(z, T, \mathfrak{R})$. While $N(z, T, \mathfrak{R})$ apparently possesses no useful continuity properties, we shall see (cf. §2.5) that *$\kappa(z, T, \mathfrak{R})$ is a lower semi-continuous function of its arguments z and T.*

1.13. Next, we consider a bounded domain(11) $\mathfrak{D}$ in the w-plane, and on it a bounded continuous transformation (cf. §1.8)

$$T: \quad z = t(w), \qquad w \in \mathfrak{D}.$$

If a sequence of Jordan regions $\mathfrak{R}_n$ is such that (i) $\mathfrak{R}_n \subset \mathfrak{D}$, and (ii) for every closed set $F \subset \mathfrak{D}$ there exists an $n_0 = n_0(F)$ such that $F \subset \mathfrak{R}_n$ for all $n > n_0$, then we shall say that the Jordan regions $\mathfrak{R}_n$ fill up $\mathfrak{D}$ from the interior. For such

(8) Since the terms *region* and *domain* do not seem to possess standardized meanings, we agree on the following terminology. A *domain* is a connected open set, while a *region* consists of a connected open set plus its boundary.

(9) For $k=1$ this concept was proposed and studied in Radó [1].

(10) Since our definition of the kernel of a finite order would remain meaningful for $k = +\infty$, it is important to note that we use an entirely different definition for the kernel of order $+\infty$.

(11) See (8).

a sequence of Jordan regions we shall see (cf. §2.6) that the sequence $\kappa(z, T, \mathfrak{R}_n)$ converges for every choice of z, the limit being possibly equal to $+\infty$. The limit is then clearly independent of the choice of the Jordan regions which fill up $\mathcal{D}$ from the interior. We define

$$\kappa(z, T, \mathcal{D}) = \lim_{n\to\infty} \kappa(z, T, \mathfrak{R}_n).$$

This function $\kappa(z, T, \mathcal{D})$—termed *the essential multiplicity of the point z in the image of $\mathcal{D}$ under T*—is fundamental in our work. To state one of our principal results, we need the following concepts.

1.14. Let $\mathfrak{R}$ be a Jordan region comprised in $\mathcal{D}$. Let C be one of the boundary curves of $\mathfrak{R}$ and let z be a point in the z-plane. As a point w describes C once in the positive sense with respect to $\mathfrak{R}$, the point $t(w)$ describes a directed closed continuous curve in the z-plane, which we shall denote by $\overline{C}$. We define $\mu(z, T, C)$ as the topological index (see Kerékjártó [1] or Radó [5]) of the point z with respect to $\overline{C}$ if z is not on $\overline{C}$; if z is on $\overline{C}$, we put $\mu(z, T, C)=0$. Finally, if z is not on the image, under T, of the boundary of $\mathfrak{R}$, we define $\mu(z, T, \mathfrak{R})=\sum\mu(z, T, C)$, the summation being extended over all the boundary curves of $\mathfrak{R}$; we set $\mu(z, T, \mathfrak{R})=0$ if z is on the image, under T, of the boundary of $\mathfrak{R}$.

1.15. If $\mu(z, T, \mathfrak{R})\neq 0$, then we shall say that $\mathfrak{R}$ is an *indicator region* for the point z under T, or briefly, an indicator region (z, T). An *indicator system* (z, T) of order k is defined as a system of k nonoverlapping indicator regions (z, T).

1.16. Given a point z in the z-plane, let us consider a component of its inverse $T^{-1}(z)$ (cf. §1.8). Such a component may or may not be a continuum (although it is always a closed set relative to $\mathcal{D}$); but if it is, we shall call it a *maximal model continuum* of z under T in $\mathcal{D}$ and we shall denote it generically by $\sigma(z, T)$. A $\sigma(z, T)$ will be called *essential* if every open set which contains it also contains a Jordan region $\mathfrak{R}$ which is an indicator region (z, T) and which contains $\sigma(z, T)$ (necessarily in its interior).

1.17. We have then the fundamental theorem([12]) (cf. §2.6):

$\kappa(z, T, \mathcal{D})$ is equal to the number of distinct essential maximal model continua of z under T in $\mathcal{D}$.

1.18. Given a bounded continuous transformation T defined on a bounded domain $\mathcal{D}$ (cf. §1.13), we define a subset $\mathcal{E}=\mathcal{E}(T, \mathcal{D})$ of $\mathcal{D}$ as follows. A point $w_0\in\mathcal{D}$ belongs to $\mathcal{E}$ if w_0 itself constitutes an essential maximal model continuum of its image $t(w_0)$. We define further a subset $\mathcal{N}=\mathcal{N}(T, \mathcal{D})$ of the set $\mathcal{E}$ as follows. A point $w_0\in\mathcal{E}$ belongs to $\mathcal{N}$ if there exists a neighborhood $\mathfrak{N}(w_0)$ of w_0 such that $\mathfrak{N}(w_0)-w_0$ contains no point of any essential maximal model continuum of $t(w_0)$.

1.19. If $w_0\in\mathcal{E}$, then we have, by definition, in every neighborhood $\mathfrak{N}(w_0)$

([12]) This theorem is one of the principal results in the dissertation cited in ([7]).

of w_0 a Jordan region $\mathfrak{R}$ such that $w_0 \in \mathfrak{R}^0$([13]) and $\mathfrak{R}$ is an indicator region $(t(w_0), T)$ (cf. §1.15). From the definition of an indicator region, it follows that $\mu(t(w_0), T, \mathfrak{R}) \neq 0$. Generally $\mu(t(w_0), T, \mathfrak{R})$ will depend upon the choice of $\mathfrak{R}$. We shall see, however, (cf. §2.20) that if $w_0 \in \mathcal{N}$, then there exists a neighborhood $\mathfrak{N}_0(w_0)$ of w_0 such that, for every choice of $\mathfrak{R}$ in $\mathfrak{N}_0(w_0)$ (as described above), the quantity $\mu(t(w_0), T, \mathfrak{R})$ has one and the same value, which we shall denote by $j(w_0, T)$, and which we shall call *the essential local index*. For w_0 not in $\mathcal{N}$, we set $j(w_0, T) = 0$. By definition then, we have

$$j(w, T) = 0 \quad \text{if} \quad w \in \mathcal{D} - \mathcal{N},$$

$$j(w, T) \neq 0 \quad \text{if} \quad w \in \mathcal{N}.$$

1.20. A fundamental fact about the essential local index $j(w, T)$ is expressed in the theorem: the set $\mathcal{N}_* = \mathcal{N}_*(T, \mathcal{D})$ of those points w where $|j(w, T)| > 1$ is always *denumerable*([14]). *It is interesting to observe that the proof of this theorem* (*cf.* §§2.31–2.40) *involves an argument whose scope seems to be restricted to the plane*—namely, the use of transformations of the form $z = w^{1/k}$, where k is a positive integer. Otherwise it seems to us that our theory applies equally well in n-dimensional Euclidean space. It would seem that a complete generalization of this theory to n-space may necessitate very fascinating geometrical investigations.

1.21. In §§1.8–1.20 we described the *topological* facts needed in our work. We turn now to its purely *metrical* aspects. Given a bounded continuous transformation T defined on a bounded domain $\mathcal{D}$ in the w-plane (cf. §1.13), let us choose a point set $\mathcal{B}$ in the w-plane (not necessarily contained in $\mathcal{D}$), which we shall term *the base set*. Concerning $\mathcal{B}$ we make the following measurability assumptions: (i) the set $\mathcal{B}$ is measurable; (ii) for every oriented([15]) rectangle R such that $R^0 \subset \mathcal{D}$([16]), the set $T(R^0 \cdot \mathcal{B})$ (cf. §1.8) is a measurable set in the z-plane.

1.22. We may now define a function of rectangles $G(R)$, for all oriented rectangles R such that $R^0 \subset \mathcal{D}$, as follows([17]):

$$G(R) = | T(R^0 \cdot \mathcal{B}) |.$$

As a matter of fact, G depends upon R, T and $\mathcal{B}$, and should accordingly be denoted by $G(R, T, \mathcal{B})$, but we shall use the notation $G(R)$ unless more explicit notation is needed.

1.23. We shall say that T is of bounded variation, with respect to the base set $\mathcal{B}$, in $\mathcal{D}$—briefly, B.V. $\mathcal{B}$ in $\mathcal{D}$—if there exists a finite constant M such that

$$\sum G(s_i) < M$$

([13]) $\mathfrak{R}^0$ denotes the set of interior points of $\mathfrak{R}$.
([14]) This theorem is a substantial generalization of a theorem in Radó [5, pp. 199–200].
([15]) See ([4]).
([16]) See ([13]).
([17]) See ([3]).

for every finite sequence of nonoverlapping, closed, oriented squares s_i such that $s_i^0 \subset \mathfrak{D}$. We shall say that T is absolutely continuous, with respect to the base set $\mathfrak{B}$ in $\mathfrak{D}$—briefly, A.C. $\mathfrak{B}$ in $\mathfrak{D}$—if, for every $\epsilon > 0$, there exists an $\eta = \eta(\epsilon) > 0$, such that

$$\sum G(s_i) < \epsilon$$

for every finite sequence of nonoverlapping, closed, oriented squares s_i such that $s_i^0 \subset \mathfrak{D}$ and

$$\sum |s_i| < \eta.$$

1.24. In the special case when the base set $\mathfrak{B}$ is the domain $\mathfrak{D}$ itself, these concepts become (essentially) identical to those of Banach (cf. §1.2). Clearly, if $\mathfrak{B}$ is a proper subset of $\mathfrak{D}$, then our definitions generally require less than those of Banach. It is a matter of fundamental importance for our work that every base set $\mathfrak{B}$, satisfying the measurability conditions stated in §1.21, gives rise to a theory as complete as that of Banach (cf. Chapter 3). We proceed to state the fundamental facts of this general theory.

1.25. Using the measurability assumptions set forth in §1.21, one sees easily (cf. §2.28) that the function $N(z, T, \mathfrak{D}\cdot\mathfrak{B})$ (cf. §1.10) is measurable. We have then the theorem[18] (cf. §3.11): *T is B.V. $\mathfrak{B}$ in $\mathfrak{D}$ if and only if $N(z, T, \mathfrak{D}\cdot\mathfrak{B})$ is a summable function.*

1.26. Given a point $w \in \mathfrak{D}$, let us consider a sequence of oriented closed squares s_n such that $w \in s_n^0 \subset \mathfrak{D}$ and $\lim |s_n| = 0$. If, for every such sequence, the quotients $G(s_n)/|s_n|$ converge to a finite limit (which is then necessarily the same for all such sequences), then this limit is called the derivative of $G(R)$ at w and will be denoted by $D(w)$—or by $D(w, T, \mathfrak{B})$ when the need for more explicit notation arises. We have then the theorem (cf. §§3.14, 3.15): *if T is B.V. $\mathfrak{B}$ in $\mathfrak{D}$, then $D(w)$ exists a.e.[19] in $\mathfrak{D}$, is measurable and summable in $\mathfrak{D}$, and we have, for every open set $O \subset \mathfrak{D}$, the inequality*[20]

$$\iint_O D(w, T, \mathfrak{B}) \leq \iint N(z, T, O\cdot\mathfrak{B}).$$

1.27. If T is B.V. $\mathfrak{B}$ in $\mathfrak{D}$, then we have, in particular,

$$\iint_{\mathfrak{D}} D(w, T, \mathfrak{B}) \leq \iint N(z, T, \mathfrak{D}\cdot\mathfrak{B}),$$

---

(18) This theorem shows that our concept of bounded variation is independent of the choice of the axes. In a similar way, further theorems in this paper can be easily shown to guarantee this invariant character for the entire theory.

(19) The abbreviation "a.e." is used consistently for "almost everywhere"—"except on a set of measure zero."

(20) For conciseness, we omit the symbols *dudv* and *dxdy* respectively in writing double integrals. In the z-plane, we shall have to integrate solely functions which vanish outside a sufficiently large disk. The range of integration will be then such a disk, and reference to it will be omitted in the formula.

and we have the theorem (cf. §§3.20, 3.21): *T is A.C. $\mathfrak{B}$ in $\mathfrak{D}$ if and only if the sign of equality holds in the preceding relation.*

1.28. Now let $H(z)$ be a finite-valued, measurable function in the z-plane. We cannot generally assert that $H(t(w))$ is measurable in $\mathfrak{D}$, but we have the theorem (cf. §3.18): if T is B.V. $\mathfrak{B}$ in $\mathfrak{D}$ and if $H(z)$ is a finite-valued, measurable function in the z-plane, then $H(t(w))D(w)$ is measurable in $\mathfrak{D}$. (Of course, this product is defined only a.e. in $\mathfrak{D}$.) Furthermore (cf. §3.19), if we are given two such functions $H_1(z)$ and $H_2(z)$, and if $H_1(z)=H_2(z)$ a.e. in the z-plane, then $H_1(t(w))D(w)=H_2(t(w))D(w)$ a.e. in $\mathfrak{D}$.

1.29. Assume that T is A.C. $\mathfrak{B}$ in $\mathfrak{D}$ (cf. §1.23), and that $H(z)$ is finite-valued and measurable in the z-plane. For every measurable set $E\subset\mathfrak{D}$ we have then *the fundamental transformation formula* (cf. §3.24):

$$\iint_E H(t(w))D(w, T, \mathfrak{B}) = \iint H(z)N(z, T, E\cdot\mathfrak{B}),$$

provided only that one of the two integrals involved exists. That is, if one of the two integrals exists, then the other one exists also, and the two are equal.

1.30. *We shall combine presently the topological facts* (cf. §§1.8–1.20) *and the metrical facts* (cf. §§1.21–1.29) *at our disposal.* Given a bounded continuous transformation T defined on a bounded domain $\mathfrak{D}$ in the w-plane (cf. §1.13), *we select the essential set $\mathcal{E}$ defined in* §1.18 *as the base set* (cf. §1.21) *in the preceding metrical theory.* We shall verify (cf. §§2.22–2.29) that $\mathcal{E}$ satisfies the measurability requirements stated in 1.21. We shall then proceed to study continuous transformations T which are B.V. $\mathcal{E}$ in $\mathfrak{D}$ and A.C. $\mathcal{E}$ in $\mathfrak{D}$ respectively (cf. §§1.23, 1.18). In the proofs (cf. Chapter 4), we shall have to consider several auxiliary set functions, together with their derivatives; the derivative $D(w, T, \mathcal{E})$ of $G(R, T, \mathcal{E})$ (cf. §1.22) will be denoted by $D_2(w)$ or by $D_2(w, T)$ when more explicit notation is needed. In case T is B.V. $\mathcal{E}$ in $\mathfrak{D}$, it follows that $D_2(w)$ exists a.e. in $\mathfrak{D}$ and is summable in $\mathfrak{D}$ (cf. §1.26). We then define a *generalized Jacobian* (cf. §4.3) by the relation

$$\mathcal{J}(w, T) = j(w, T)D_2(w, T)$$

which we also denote by $\mathcal{J}(w)$. From §1.20 it follows that $|\mathcal{J}(w)| \leq D_2(w)$ a.e. in $\mathfrak{D}$; hence $\mathcal{J}(w)$ is also summable in $\mathfrak{D}$.

1.31. Given any domain $\mathfrak{D}_*\subset\mathfrak{D}$, we define an *essential index function* $\nu(z, T, \mathfrak{D}_*)$ as follows (cf. §§1.8, 1.10, 1.18, 1.19):

$$\nu(z, T, \mathfrak{D}_*) = \begin{cases} \sum_w j(w, T) \text{ for } w \in \mathcal{E}\cdot\mathfrak{D}_*\cdot T^{-1}(z), \text{ if } z \text{ not in } \overline{\mathfrak{S}}(\infty, T, \mathcal{E}\cdot\mathfrak{D}_*), \\ 0 \text{ if } z \in \overline{\mathfrak{S}}(\infty, T, \mathcal{E}\cdot\mathfrak{D}_*). \end{cases}$$

We shall show (cf. §2.30) that $\nu(z, T, \mathfrak{D}_*)$ is a summable function provided that T is B.V. $\mathcal{E}$ in $\mathfrak{D}_*$.

1.32. *Given a bounded domain $\mathfrak{D}$ in the w-plane, our fundamental class of transformations on $\mathfrak{D}$ will be denoted by $K_1(\mathfrak{D})$ and will consist of all those bounded continuous transformations defined on $\mathfrak{D}$* (cf. §1.13) *which are A.C. $\mathcal{E}$ in $\mathfrak{D}$*([21]) (cf. §1.23). By the general metric theory described above (cf. §1.29), we have then, for every finite-valued, measurable function $H(z)$ and for every transformation $T \in K_1(\mathfrak{D})$ *the metrical transformation formula*

$$\iint_{\mathfrak{D}} H(t(w))D_2(w, T) = \iint H(z)N(z, T, \mathfrak{D}\cdot\mathcal{E}),$$

as soon as one of these integrals exists. We shall establish (cf. §4.9), for every finite-valued, measurable function $H(z)$ and for every transformation $T \in K_1(\mathfrak{D})$ *the topological transformation formula* (cf. §1.31)

$$\iint_{\mathfrak{D}} H(t(w))\mathcal{J}(w, T) = \iint H(z)\nu(z, T, \mathfrak{D}),$$

provided that the integral on the left exists. That is, if the integral on the left exists, then the integral on the right exists, and the integrals are equal. The converse is generally false (see Radó [5, p. 214]).

1.33. The preceding formulas are of extreme generality. In particular, *the topological transformation formula of* §1.32 *implies all the results on the transformation of double integrals in the literature of which we are aware* (cf. Bibliography). As a matter of fact, very special cases of the preceding results are ample for that purpose. We proceed now to describe such special cases for the purpose of applications to be presented on another occasion.

1.34. Given a bounded domain $\mathfrak{D}$ in the w-plane, let $K_2(\mathfrak{D})$ denote the class of all transformations $T \in K_1(\mathfrak{D})$ (cf. §1.32) for which $N(z, T, \mathcal{E}\cdot\mathfrak{D}) = \kappa(z, T, \mathfrak{D})$ a.e. in the z-plane (cf. §§1.10, 1.13, 1.18). *For transformations $T \in K_2(\mathfrak{D})$ we shall see* (cf. §4.7) *that $|\mathcal{J}(w)| = D_2(w)$ a.e. in $\mathfrak{D}$* (cf. §1.30). Next, we shall consider the special case when $\mathfrak{D}$ is the interior $\mathfrak{R}^0$ of a Jordan region $\mathfrak{R}$ (cf. §4.9). For this case, we shall derive the following theorem: *if T is continuous on the Jordan region $\mathfrak{R}$, if $T \in K_2(\mathfrak{R}^0)$, and if the image of the boundary of $\mathfrak{R}$ is a set of measure zero, then, for every finite-valued, measurable function $H(z)$ we have* (cf. §1.14)

$$\iint_{\mathfrak{R}^0} H(t(w))\mathcal{J}(w) = \iint H(z)\mu(z, T, \mathfrak{R}),$$

as soon as the integral on the left exists.

1.35. A fundamental property of the class $K_2(\mathfrak{D})$ is expressed in the following:

([21]) That is, $T \in K_1(\mathfrak{D})$ if T is absolutely continuous in $\mathfrak{D}$ with respect to its own essential set $\mathcal{E} = \mathcal{E}(T, \mathfrak{D})$.

Closure theorem. *Let there be given bounded domains* $\mathcal{D}$ *and* $\mathcal{D}_n$ *in the w-plane, and continuous transformations* (*cf.* §1.8)

$$T: \quad z = t(w),\ w \in \mathcal{D}, \qquad T_n: \quad z = t_n(w),\ w \in \mathcal{D}_n,$$

with the following properties: (i) *the domains* $\mathcal{D}_n$ *fill up* $\mathcal{D}$ *from the interior*([22]); (ii) *the generalized Jacobian* $\mathcal{J}(w, T)$ *exists a.e. in* $\mathcal{D}$ *and is summable in* $\mathcal{D}$; (iii) $T_n \in K_2(\mathcal{D}_n)$ *for* $n = 1, 2, \cdots$; (iv) *for every closed set* $F \subset \mathcal{D}$ *we have* $\lim \rho(T_n, T; F) = 0$ (*cf.* §1.9), *and*

$$\lim \iint_F \left| \mathcal{J}(w, T) - \mathcal{J}(w, T_n) \right| = 0.$$

Then $T \in K_2(\mathcal{D})$ (*cf.* §4.11).

1.36. Next, we have the following theorem (cf. §4.14): in a Jordan region $\mathfrak{R}$ let there be given continuous transformations T and T_n with the following properties: (i) $T \in K_2(\mathfrak{R}^0)$, $T_n \in K_2(\mathfrak{R}^0)$; (ii) $\lim \rho(T_n, T; \mathfrak{R}) = 0$; (iii) the image of the boundary of $\mathfrak{R}$ under T is a set of measure zero; (iv) we have

$$\lim \iint_{\mathfrak{R}^0} \left| \mathcal{J}(w, T) - \mathcal{J}(w, T_n) \right| = 0.$$

Then (cf. §1.14)

$$\lim \iint \left| \mu(z, T, \mathfrak{R}) - \mu(z, T_n, \mathfrak{R}) \right| = 0.$$

That is, *strong convergence of the generalized Jacobians* (cf. §1.30) *implies strong convergence of the index functions* μ (cf. §1.14).

1.37. We now write the transformation T in the form (cf. §1.8)

$$T: \quad x = x(u, v), \qquad y = y(u, v), \qquad (u, v) \in \mathcal{D},$$

and denote by $K_3(\mathcal{D})$ the class of all transformations $T \in K_2(\mathcal{D})$ (cf. §1.34) for which the ordinary Jacobian $J(w, T) = J(u, v, T) = x_u y_v - x_v y_u$ exists a.e. in $\mathcal{D}$.

Note that this is the first time that we refer to the ordinary Jacobian, either in the theorems or in the proofs.

1.38. We have the important theorem (cf. §5.5): *if* $T \in K_3(\mathcal{D})$, *then* $J(w, T) = \mathcal{J}(w, T)$ *a.e. in* $\mathcal{D}$. This theorem permits us to replace $\mathcal{J}(w, T)$ by $J(w, T)$ in the transformation formulas in §1.32 and §1.34. *In this fashion we obtain formulas which include all similar results in the literature of which we are aware.* We now list some theorems which account for these results.

1.39. Suppose that the functions $x(u, v)$, $y(u, v)$ defining T (cf. §1.37) satisfy a Lipschitz condition in $\mathcal{D}$ in the following restricted sense: there exists

([22]) The domains $\mathcal{D}_n$ are said to fill up $\mathcal{D}$ from the interior if (i) $\mathcal{D}_n \subset \mathcal{D}$ and (ii) for every closed set $F \subset \mathcal{D}$, there exists an $n_0 = n_0(F)$ such that $F \subset \mathcal{D}_n$ for all $n > n_0$.

a finite constant L such that, if (u_1, v_1), (u_2, v_2) are any two points in $\mathfrak{D}$ for which the line segment joining them is contained in $\mathfrak{D}$, then

$$\left.\begin{array}{l} | x(u_2, v_2) - x(u_1, v_1) | \\ | y(u_2, v_2) - y(u_1, v_1) | \end{array}\right\} \leqq L[(u_2 - u_1)^2 + (v_2 - v_1)^2]^{1/2}.$$

Then $T \in K_3(\mathfrak{D})$ (cf. §5.8).

1.40. Suppose that the partial derivatives x_u, x_v, y_u, y_v exist and are continuous in $\mathfrak{D}$ and that the ordinary Jacobian $J(w, T)$ is summable in $\mathfrak{D}$. Then $T \in K_3(\mathfrak{D})$ (cf. §5.9).

1.41. For the class $K_3(\mathfrak{D})$ we have the following:

Closure theorem. *Let there be given bounded domains $\mathfrak{D}$ and $\mathfrak{D}_n$ in the uv-plane, and continuous transformations* (*cf.* §1.8)

$$T: \quad x = x(u, v), \qquad y = y(u, v), \qquad (u, v) \in \mathfrak{D},$$

$$T_n: \quad x = x_n(u, v), \qquad y = y_n(u, v), \qquad (u, v) \in \mathfrak{D}_n,$$

with the following properties: (i) *the domains $\mathfrak{D}_n$ fill up $\mathfrak{D}$ from the interior*[23]; (ii) *the ordinary Jacobian $J(w, T)$ exists a.e. in $\mathfrak{D}$ and is summable in $\mathfrak{D}$*; (iii) $T_n \in K_3(\mathfrak{D}_n)$ *for* $n = 1, 2, \cdots$; (iv) *for every closed set $F \subset \mathfrak{D}$ we have* $\lim \rho(T_n, T; F) = 0$ (*cf.* §1.9), *and*

$$\lim \iint_F | J(w, T) - J(w, T_n) | = 0.$$

Then $T \in K_3(\mathfrak{D})$ (*cf.* §5.7).

1.42. The reader will note that the ordinary Jacobian $J(u, v) = x_u y_v - x_v y_u$ appears only in the last few sections, *after* the general theory has been completed. *As far as our theory is concerned, the ordinary Jacobian is entirely irrelevant, and we consider it solely for the purpose of applications to previous literature.* This situation leads quite naturally to the question as to whether the ordinary Jacobian is really an essential concept in formulating and studying fundamental problems involving double integrals in analysis and in geometry.

Chapter II. Topological foundations

2.1. In this chapter we shall give concise proofs of the geometrical facts in this theory. Most of these facts have been stated in §§1.8–1.20. In order to verify them, however, we shall find it necessary to use many auxiliary results; for the proofs of these, we refer the reader to the literature, unless we are aware of no simple published proofs.

2.2. Let $\mathfrak{R}$ be a Jordan region in the w-plane (cf. §1.11), and let T be a

(23) See (22).

continuous transformation defined on $\mathfrak{R}$ (cf. 1.8). We state two properties of the kernel of order k (cf. §1.11) needed in the sequel.

Lemma 1. *Every kernel $\overline{\mathfrak{K}}(k, T, \mathfrak{R})$ of finite order is an open set*[24].

Proof[25]. If $\overline{\mathfrak{K}}$ is empty, the lemma is trivial. So suppose $z \in \overline{\mathfrak{K}}$; we assert that every point z_* satisfying $|z_* - z| < \epsilon = \epsilon(k, z, T, \mathfrak{R})$ (cf. §1.11) is also in $\overline{\mathfrak{K}}$. Set $\lambda = z - z_*$, $\epsilon_* = \epsilon - |\lambda|$; clearly $\epsilon_* > 0$. If T_* be any continuous transformation defined on $\mathfrak{R}$ and satisfying $\rho(T_*, T; \mathfrak{R}) < \epsilon_*$ (cf. 1.9), then clearly the transformation T^* defined by $t^*(w) = t_*(w) + \lambda$, $w \in \mathfrak{R}$, is continuous on $\mathfrak{R}$ and satisfies $\rho(T^*, T; \mathfrak{R}) < \epsilon$; consequently $N(z, T^*, \mathfrak{R}) \geqq k$ (cf. §1.11). But obviously $N(z_*, T_*, \mathfrak{R}) = N(z, T^*, \mathfrak{R}) \geqq k$; thus $z^* \in \overline{\mathfrak{K}}$ by definition.

Lemma 2. *If* $\{T_n\}$ *be a sequence of continuous transformations defined on* $\mathfrak{R}$ *and satisfying* $\lim \rho(T_n, T; \mathfrak{R}) = 0$, *then* $\liminf \overline{\mathfrak{K}}(k, T_n, \mathfrak{R}) \supset \overline{\mathfrak{K}}(k, T, \mathfrak{R})$, *provided k is finite.*

The proof is obvious (see Radó [5, p. 205]).

2.3. We pause to introduce some definitions and notations useful in the sequel. Given a function $f(w)$ defined on some set in the w-plane, and any set E on which it is defined, we denote by $M(f, E)$ the least upper bound of $|f(w)|$ for $w \in E$; we denote by $O(f, E)$ the least upper bound of $|f(w') - f(w'')|$ for $w', w'' \in E$. Obviously $O(f, E) \leqq 2M(f, E)$. If δ be any positive number, we denote by $O(\delta, f, E)$ the least upper bound of $|f(w') - f(w'')|$ for $w', w'' \in E$, $|w' - w''| \leqq \delta$. If $f(w)$ is continuous and E is a bounded closed set, then $O(\delta, f, E)$ converges to zero with δ.

2.4. In §1.12 we observed that $\kappa(z, T, \mathfrak{R}) \leqq N(z, T, \mathfrak{R})$. However, we shall have use for the stronger result contained in the

Lemma. $\kappa(z, T, \mathfrak{R}) \leqq N(z, T, \mathfrak{R}^0)$[26].

Proof. Clearly it is sufficient to show that, if $N(z, T, \mathfrak{R}^0) = k < +\infty$, then, for every $\epsilon > 0$, there exists a continuous transformation T_ϵ defined on $\mathfrak{R}$ and satisfying $\rho(T_\epsilon, T; \mathfrak{R}) < \epsilon$ and $N(z, T_\epsilon, \mathfrak{R}) \leqq k$. Let $\overline{\mathfrak{R}}$ be a Jordan region (cf. §1.11) in a $\bar{w}$-plane having the same connectivity as $\mathfrak{R}$, each of whose boundary curves is a circle. Map $\mathfrak{R}$ onto $\overline{\mathfrak{R}}$ by a biunique and continuous transformation given by $\bar{w} = f(w)$, $w \in \mathfrak{R}$ (see Kerékjártó [1]). The inverse transformation given by $w = f^{-1}(\bar{w})$, $\bar{w} \in \overline{\mathfrak{R}}$, is also biunique and continuous, and maps $\overline{\mathfrak{R}}$ onto $\mathfrak{R}$. Define a continuous transformation $\overline{T}$ from $\overline{\mathfrak{R}}$ to the z-plane by the relation $z = \bar{t}(\bar{w}) = t(f^{-1}(\bar{w}))$, $\bar{w} \in \overline{\mathfrak{R}}$ (cf. §1.8). Let δ be a positive number such that $O(\delta, \bar{t}, \overline{\mathfrak{R}}) < \epsilon$ (cf. §2.3). If $\overline{C}$ denotes a typical boundary circle of $\overline{\mathfrak{R}}$, choose in $\overline{\mathfrak{R}}^0$ a concentric circle $\overline{C}_\epsilon$ at a distance less than δ from $\overline{C}$ so that the

(24) Radó (see Radó [5, p. 203]) proved the weaker result that every kernel of finite order is an F_σ.

(25) For brevity, we write $\overline{\mathfrak{K}}$ for $\overline{\mathfrak{K}}(k, T, \mathfrak{R})$ if the meaning is clear from the context.

(26) See (19).

circles $\bar{C}_\epsilon$ bound a Jordan region $\bar{\mathfrak{R}}_\epsilon$ in $\bar{\mathfrak{R}}^0$ having the same connectivity as $\bar{\mathfrak{R}}$ and for which $N(z, \bar{T}, \bar{\mathfrak{R}}_\epsilon) = N(z, \bar{T}, \bar{\mathfrak{R}}_\epsilon^0) \leqq k$. Define a continuous transformation $\bar{T}_\epsilon$ from $\mathfrak{R}$ to the z-plane by the relations

$$\bar{T}_\epsilon: \quad z = \bar{t}_\epsilon(\bar{w}) = \begin{cases} \bar{t}(\bar{w}), & \bar{w} \in \bar{\mathfrak{R}}_\epsilon; \\ \bar{t}(\bar{w}_*), & \bar{w}_* \in \bar{\mathfrak{R}} - \bar{\mathfrak{R}}_\epsilon, \end{cases}$$

where $\bar{w}_*$ is the closest point to $\bar{w}$ on a boundary curve of $\bar{\mathfrak{R}}_\epsilon$. Clearly $\rho(\bar{T}_\epsilon, \bar{T}; \bar{\mathfrak{R}}) < \epsilon$ and $N(z, \bar{T}_\epsilon, \bar{\mathfrak{R}}) = N(z, \bar{T}, \bar{\mathfrak{R}}_\epsilon^0) \leqq k$. Finally, set T_ϵ: $z = t_\epsilon(w) = \bar{t}_\epsilon(f(w))$, $w \in \mathfrak{R}$. Clearly the transformation T_ϵ thus defined (cf. §1.8) has the required properties.

2.5. As a corollary to the results in §2.2 and §2.4 and to the definition of the essential multiplicity (cf. §1.12), it follows that $\kappa(z, T, \mathfrak{R})$ is a lower semi-continuous function of z and of T[27], as asserted in §1.12.

2.6. We turn now to establish the important

Theorem. *$\kappa(z, T, \mathfrak{R})$ is equal to the number of distinct essential maximal model continua of z under T in $\mathfrak{R}^0$ (cf. §1.16).*

The proof of this theorem will follow (cf. §2.17) from a series of simple facts which we presently state and verify. The reader will easily see that this theorem accounts for all the assertions made in §1.13 and for the fundamental characterization theorem for the essential multiplicity function $\kappa(z, T, \mathfrak{D})$ for a domain $\mathfrak{D}$ which is stated in §1.17.

2.7. To approach the proof of the theorem stated in §2.6 (and for later purposes also), we need the following facts concerning the function $\mu(z, T, \mathfrak{R})$ defined in §1.14. Given a point z_0 in the z-plane, define $\epsilon(z_0, T, \mathfrak{R})$ to be the minimum of $|t(w) - z_0|$ for $w \in \mathfrak{R} - \mathfrak{R}^0$[28]. Clearly $\epsilon(z_0, T, \mathfrak{R}) = 0$ if and only if $z_0 \in T(\mathfrak{R} - \mathfrak{R}^0)$ (cf. §1.8).

Lemma 1. *Consider the continuous transformations (cf. §1.8)*

$$T_i: \quad z = t_i(w), \qquad w \in \mathfrak{R},\ i = 0, 1, \cdots, m.$$

Suppose that, for points $z_0, z_1, \cdots, z_m$, it is true that $\epsilon(z_i, T_i, \mathfrak{R}) > 0$ for $i = 1, \cdots, m$, and

$$t_0(w) - z_0 = \prod_{i=1}^{m} [t_i(w) - z_i], \qquad w \in \mathfrak{R}.$$

Then

$$\mu(z_0, T_0, \mathfrak{R}) = \sum_{i=1}^{m} \mu(z_i, T_i, \mathfrak{R}).$$

---

[27] A sequence of continuous transformations T_n defined on $\mathfrak{R}$ is said to converge to T on $\mathfrak{R}$ if $\lim \rho(T_n, T; \mathfrak{R}) = 0$ (cf. §1.9).

[28] $\mathfrak{R} - \mathfrak{R}^0$ is the set of boundary points of $\mathfrak{R}$. See [8], [13].

This lemma is a ready consequence of the properties of the topological index (cf. §1.14).

LEMMA 2. *Assume that* $\epsilon(z, T, \mathfrak{R})>0$. ($\alpha$) *If* z_* *be any point in the z-plane satisfying* $|z_*-z|<\epsilon(z, T, \mathfrak{R})$, *then* $\mu(z_*, T, \mathfrak{R})=\mu(z, T, \mathfrak{R})$. ($\beta$) *If* T_* *be any continuous transformation defined on* $\mathfrak{R}$ *and satisfying* $\rho(T_*, T; \mathfrak{R}-\mathfrak{R}^0)<\epsilon(z, T, \mathfrak{R})$, *then* $\mu(z, T_*, \mathfrak{R})=\mu(z, T, \mathfrak{R})$.

This lemma is an immediate consequence of the *theorem of Rouché*[29].

LEMMA 3. $\mu(z, T, \mathfrak{R})=0$ *for* z *not* $\in T(\mathfrak{R})$ (see Radó [5, p. 196] also for further references).

LEMMA 4. *Assume that* $\mathfrak{R}$ *is simply connected. Let* z_0 *be a point of the z-plane for which* $\epsilon(z_0, T, \mathfrak{R})>0$ *and* $\mu(z_0, T, \mathfrak{R})=0$. *Then there exists a function* $f(w)$ *having the following properties*: (i) $f(w)$ *is defined and continuous on* $\mathfrak{R}$; (ii) $f(w)=t(w)$ *for* $w\in\mathfrak{R}-\mathfrak{R}^0$; (iii) $M(f(w)-z_0, \mathfrak{R})=M(t(w)-z_0, \mathfrak{R}-\mathfrak{R}^0)$; (iv) $f(w)\neq z_0$ *for* $w\in\mathfrak{R}$ (see Radó [5, p. 196] also for further references).

LEMMA 5. *Let* $\mathfrak{R}_1, \cdots, \mathfrak{R}_m$ *be any finite number of nonoverlapping Jordan regions in* $\mathfrak{R}-\mathfrak{R}_0$. *Given a point* z_0 *in the z-plane, let* $\zeta=\zeta(z_0, T, \{\mathfrak{R}_i\}_0^m)$ *denote the greatest lower bound of* $|t(w)-z_0|$ *for* $w\in\mathfrak{R}_0-\sum_{i=1}^m\mathfrak{R}_i^0$. *If* z_0 *be any point for which* $\zeta>0$, *then* $\mu(z_0, T, \mathfrak{R}_0)=\sum_{i=1}^m\mu(z_0, T, \mathfrak{R}_i)$.

Proof. Clearly $\epsilon(z_0, T, \mathfrak{R}_i)\geqq\zeta>0$ for $i=0, 1, \cdots, m$. Let T_* be any continuous transformation defined on $\mathfrak{R}$, and such that $\rho(T_*, T; \mathfrak{R})<\zeta$ and z_0 not $\in\mathfrak{S}(\infty, T_*, \mathfrak{R})$ (cf. §1.10)[30]. From Lemma 2, it follows that $\mu(z_0, T_*, \mathfrak{R}_i)=\mu(z_0, T, \mathfrak{R}_i)$ for $i=0, 1, \cdots, m$. Since z_0 not $\in\mathfrak{S}(\infty, T_*, \mathfrak{R})$, the sets $T_*^{-1}(z_0)\cdot\mathfrak{R}_i^0$ are finite for $i=0, 1, \cdots, m$ (cf. §§1.8, 1.10). We shall discuss the general case when none of these sets are empty, leaving special cases for the reader. If $w_{i1}, \cdots, w_{ij_i}$ denote the points of $T_*^{-1}(z_0)\cdot\mathfrak{R}_i^0$ for $i=1, \cdots, m$, then clearly the points in the set $T_*^{-1}(z_0)\cdot\mathfrak{R}_0^0$ consist of

$$w_{11}, \cdots, w_{1j_1}, \cdots, w_{m1}, \cdots, w_{mj_m}.$$

Choose j_i mutually exclusive simply connected Jordan regions $\mathfrak{R}_{ij}$ such that $w_{ij}\in\mathfrak{R}_{ij}^0$ for $j=1, \cdots, j_i$, $i=1, \cdots, m$. As a consequence of Lemma 3, and of the assumption $\zeta>0$, it follows that $\mu(z_0, T_*, \mathfrak{R}_i)=\sum_{j=1}^{j_i}\mu(z_0, T_*, \mathfrak{R}_{ij})$ for $i=1, 2, \cdots, m$, and

(29) We mean the purely topological statement of the theorem of Rouché as given, for example, in Szilárd [1, p. 655].

(30) The existence of such a transformation as T_* may be inferred as follows. Since every Jordan region is topologically equivalent to a polygonal region of the same connectivity, we may assume that $\mathfrak{R}$ is a polygonal region—that is, a Jordan region each of whose boundary curves is a simple polygon. Given $\zeta>0$ one may choose a triangulation of $\mathfrak{R}$ so fine that one may define upon $\mathfrak{R}$ a continuous transformation T_ζ which is linear and biunique in each triangle of the triangulation, and for which $\rho(T_\zeta, T; \mathfrak{R})<\zeta$. Clearly $\mathfrak{S}(\infty, T_\zeta, \mathfrak{R})$ is empty; thus T_ζ has the desired properties.

$$\mu(z_0, T_*, \mathfrak{R}_0) = \sum_{i=1}^{m} \sum_{j=1}^{j_i} \mu(z_0, T_*, \mathfrak{R}_{ij}) = \sum_{i=1}^{m} \mu(z_0, T_*, \mathfrak{R}_i).$$

In view of preceding equalities, the lemma now follows.

2.8. Using Lemma 2 in §2.7 and the definition of an indicator system (z, T) (cf. §1.15), the reader will easily verify the

LEMMA. *If $\mathfrak{S}$ be any indicator system (z, T) of order k, then $\epsilon(z, T, \mathfrak{S})=$ the minimum of $\epsilon(z, T, \mathfrak{R}_*)$ for $\mathfrak{R}_* \in \mathfrak{S}$, is positive. ($\alpha$) For every point z_* satisfying $|z_* - z| < \epsilon(z, T, \mathfrak{S})$, it is true that $\mathfrak{S}$ is also an indicator system (z_*, T) of order k, and $\mu(z_*, T, \mathfrak{R}_*) = \mu(z, T, \mathfrak{R}_*)$ for every $\mathfrak{R}_* \in \mathfrak{S}$. ($\beta$) For every continuous transformation T_* defined on $\mathfrak{R}$ and satisfying $\rho(T_*, T; \mathfrak{R}_* - \mathfrak{R}_*^0) < \epsilon(z, T, \mathfrak{S})$ for every $\mathfrak{R}_* \in \mathfrak{S}$, it is true that $\mathfrak{S}$ is also an indicator system (z, T_*) of order k, and $\mu(z, T_*, \mathfrak{R}_*) = \mu(z, T, \mathfrak{R}_*)$ for every $\mathfrak{R}_* \in \mathfrak{S}$.*

2.9. Let z_0 be any point in the z-plane. The least upper bound of the orders of all indicator systems (z_0, T) in $\mathfrak{R}^0$ will be denoted by $\psi(z_0, T, \mathfrak{R}^0)$. It follows from Lemma 3 in §2.7 that every indicator region (z_0, T) contains (in its interior) at least one point of $T^{-1}(z_0)$ (cf. §1.15). Consequently, in view of the definitions of $\kappa(z_0, T, \mathfrak{R})$ (cf. §1.12), and of an indicator system (z_0, T) (cf. §1.15), it follows from the lemma in §2.8 that $\psi(z_0, T, \mathfrak{R}^0) \leqq \kappa(z_0, T, \mathfrak{R})$. Further, if $\phi(z_0, T, \mathfrak{R}^0)$ denotes the number of essential maximal model continua of z_0 under T in $\mathfrak{R}^0$ (cf. §1.16), it is clear that $\phi(z_0, T, \mathfrak{R}^0) \leqq \psi(z_0, T, \mathfrak{R}^0)$.

2.10. Assume, for the moment, that z not $\in \overline{\mathfrak{S}}(\infty, T, \mathfrak{R}^0)$ (cf. §1.10). Then the point set $T^{-1}(z) \cdot \mathfrak{R}^0$ is finite; consequently, if $w \in T^{-1}(z) \cdot \mathfrak{R}^0$, then there exists a neighborhood $\mathfrak{N}_0(w)$ of w containing no further points of $T^{-1}(z)$. Now, for every Jordan region $\mathfrak{R}_*$ in $\mathfrak{N}_0(w)$ and containing w in its interior, it follows by Lemma 3 in §2.7 that $\mu(t(w), T, \mathfrak{R}_*)$ has the same value. If $\mu(t(w), T, \mathfrak{R}_*) \neq 0$, then clearly w constitutes an essential maximal model continuum of $z = t(w)$ belonging to the set $\mathcal{N}(T, \mathfrak{R}^0)$ (cf. §1.18). Thus the assertion made in §1.19 regarding points of $\mathcal{N}$ is now verified for all points $w \in \mathcal{N}$ for which $t(w)$ is not $\in \overline{\mathfrak{S}}(\infty, T, \mathfrak{R}^0)$. Further, it is obvious that $\phi(z, T, \mathfrak{R}^0) = N(z, T, \mathcal{N})$ in this case (cf. §2.9). But, moreover, we have the

LEMMA. *If z not $\in \overline{\mathfrak{S}}(\infty, T, \mathfrak{R}^0)$, then $\kappa(z, T, \mathfrak{R}) = N(z, T, \mathcal{N})$*[31].

Proof. A point z_0 not $\in \overline{\mathfrak{S}}(\infty, T, \mathfrak{R}^0)$ has a finite number of models under T in $\mathfrak{R}^0$; denote the points of $T^{-1}(z_0)$ in $\mathcal{N}$ by $w_1, \cdots, w_k$; denote the points of $T^{-1}(z_0)$ in $\mathfrak{R}^0 - \mathcal{N}$ by $w_{k+1}, \cdots, w_l$. We shall first show that $\kappa \geqq k$[32]. Evidently one may choose k mutually exclusive (simply connected) Jordan regions $\mathfrak{R}_j \subset \mathfrak{R}^0$ so that $w_j \in \mathfrak{R}_j^0$ and $T^{-1}(z_0) \cdot \mathfrak{R}_j = w_j$ for $j = 1, \cdots, k$. Since

---

(31) Radó (see Radó [5, p. 204]) used the same methods in proving a slightly less general result.

(32) For brevity, we write κ for $\kappa(z, T, \mathfrak{R})$ where no misunderstanding is possible.

clearly $\mu(z_0, T, \mathfrak{R}_j)=j(w_j, T)\neq 0$ for $j=1, \cdots, k$ (cf. §1.19), it follows by §1.15 that the k Jordan regions $\mathfrak{R}_j$ constitute an indicator system (z_0, T) of order k, hence $\kappa \geqq k$ (cf. §2.9). We shall next prove that $\kappa \leqq k$. In view of the lemma in §2.4, it is sufficient to exhibit, for every $\epsilon>0$, a continuous transformation T_ϵ defined on $\mathfrak{R}$ for which $\rho(T_\epsilon, T; \mathfrak{R})<\epsilon$ and $N(z_0, T_\epsilon, \mathfrak{R}^0)\leqq k$. One may obviously choose $l-k$ mutually exclusive simply connected Jordan regions $\mathfrak{R}_j \subset \mathfrak{R}^0$ so that $w_j \in \mathfrak{R}_j^0$, $T^{-1}(z_0)\cdot \mathfrak{R}_j = w_j$, and $M(t(w)-z_0, \mathfrak{R}_j)<\epsilon/2$ for $j=k+1, \cdots, l$. Since $w_j \in \mathfrak{R}^0-\mathfrak{N}$, it follows that $\mu(z_0, T, \mathfrak{R}_j)=j(w_j, T)=0$ for $j=k+1, \cdots, l$ (cf. §1.19). Thus by Lemma 4 in §2.7, there exist $l-k$ functions $f_j(w)$ for $j=k+1, \cdots, l$ with the following properties: (i) $f_j(w)$ is defined and continuous on $\mathfrak{R}_j$; (ii) $f_j(w)=t(w)$ for $w\in \mathfrak{R}_j-\mathfrak{R}_j^0$; (iii) $M(f_j(w)-z_0, \mathfrak{R}_j)<\epsilon/2$; (iv) $f_j(w_0)\neq z_0$ for $w\in\mathfrak{R}_j$. Define T_ϵ by the relations

$$T_\epsilon:\quad z = t_\epsilon(w) = \begin{cases} t(w) & \text{if } w \in \mathfrak{R} - \sum_{j=k+1}^{l} \mathfrak{R}_j, \\ f_j(w) & \text{if } w \in \mathfrak{R}_j, \quad j = k+1, \cdots, l. \end{cases}$$

Clearly T_ϵ is a continuous transformation defined on $\mathfrak{R}$, $\rho(T_\epsilon, T; \mathfrak{R})<\epsilon$, and $N(z_0, T_\epsilon, \mathfrak{R}^0)=k$. Combining preceding inequalities, we have $\kappa(z_0, T, \mathfrak{R})=k$ $=N(z_0, T, \mathfrak{N})$, as asserted.

2.11. Surveying the remarks in §§2.9, 2.10 and the lemma in §2.10, we observe that the theorem stated in §2.6 is now established in the special case when z not $\in \overline{\mathfrak{S}}(\infty, T, \mathfrak{R}^0)$.

2.12. Given a continuous transformation T defined on a Jordan region $\mathfrak{R}$ (cf. §1.8). Let τ be any positive number, and z_0 any point of the z-plane; denote by $G=G(\tau, z_0, T, \mathfrak{R})$ the set of all points $w\in\mathfrak{R}^0$ for which $|t(w)-z_0|<\tau$. Clearly $T^{-1}(z_0)\subset G$. Since G is an open set, it is the sum of at most a denumerable number of mutually exclusive components, which we shall denote generically by $g=g(\tau, z_0, T, \mathfrak{R})$. Each set g is a connected open set; denote by $\bar{g}$ the set g plus all its boundary points. Clearly $|t(w)-z_0|=\tau$ for $w\in(\bar{g}-g)\cdot\mathfrak{R}^0$; thus, if w_0 be a point of $\bar{g}-g$ for which $|t(w_0)-z_0|<\tau$, it follows that $w_0\in\mathfrak{R}-\mathfrak{R}^0$.

2.13. Lemma 1. *Assume that z_0 not $\in\overline{\mathfrak{S}}(\infty, T, \mathfrak{R}^0)$. If g be any component of $G=G(\tau, z_0, T, \mathfrak{R})$ for which $|t(w)-z_0|=\tau$ for $w\in\bar{g}-g$ and $N(z_0, T, g)>0$, then there exists a simply connected Jordan region $\mathfrak{R}_*\subset g$ and a continuous transformation T_* defined on $\mathfrak{R}_*$ and satisfying the following conditions*: (i) $\rho(T_*, T; \mathfrak{R}_*-\mathfrak{R}_*^0)=0$; (ii) $\rho(T_*, T; \mathfrak{R}_*)<2\tau$; (iii) $N(z_0, T, g-\mathfrak{R}_*^0)=0$; (iv) $N(z_0, T_*, \mathfrak{R}_*)$ $=N(z_0, T_*, \mathfrak{R}_*^0)=1$.

Proof. Our assumptions imply that there exists a simply connected Jordan region $\mathfrak{R}_*\subset g$ which contains the set $T^{-1}(z_0)\cdot g$ in its interior. Map $\mathfrak{R}_*$ onto the closed unit disk $\overline{\mathfrak{R}}_*$ with center at the origin in the $\bar{w}$-plane by a biunique and continuous transformation given by $\bar{w}=f(w)$, $w\in\mathfrak{R}_*$ (cf. Kerékjártó [1]).

Define

$$\bar{t}_*(\bar{w}) = \begin{cases} t(f^{-1}(\bar{w})) \text{ if } w \in \overline{\mathfrak{R}}_* - \overline{\mathfrak{R}}_*^0, \\ z_0 \text{ if } \bar{w} = 0, \\ (\bar{t}_*(\bar{w}_*) - z_0) \mid \bar{w} \mid + z_0 \text{ if } \bar{w} \in \overline{\mathfrak{R}}_*^0 - 0, \end{cases}$$

where $\bar{w}_*$ is the unique point of $\overline{\mathfrak{R}}_* - \overline{\mathfrak{R}}_*^0$ on the ray from 0 through $\bar{w}$. Finally, set T_*: $z = t_*(w) = \bar{t}_*(f(w))$, $w \in \mathfrak{R}_*$. Clearly the Jordan region $\mathfrak{R}_*$ and the transformation T_* thus defined have the required properties.

Lemma 2. *Assume that z_0 not $\in \mathfrak{S}(\infty, T, \mathfrak{R}^0)$. If g be any component of $G = G(\tau, z_0, T, \mathfrak{R})$ for which there exists a point $w_0 \in \bar{g} - g$ such that $|t(w_0) - z_0| < \tau$ and $N(z_0, T, g) > 0$, then there exists a simply connected Jordan region $\mathfrak{R}_* \subset \bar{g}$ and a continuous transformation T_* defined on $\mathfrak{R}_*$ and satisfying the following conditions*: (i) $\rho(T_*, T; (\mathfrak{R}_* - \mathfrak{R}_*^0) \cdot \mathfrak{R}^0) = 0$; (ii) $\rho(T_*, T; \mathfrak{R}_*) < 2\tau$; (iii) $N(z_0, T, g - \mathfrak{R}_*^0) = 0$; (iv) $N(z_0, T_*, \mathfrak{R}_* \cdot \mathfrak{R}^0) = 0$.

Using the fact that $w_0 \in \mathfrak{R} - \mathfrak{R}^0$ (cf. §2.12), the reader may construct the proof by methods similar to those used in the proof of Lemma 1.

2.14. Lemma. *$\psi(z, T, \mathfrak{R}^0) \geqq \kappa(z, T, \mathfrak{R})$ (cf. §§1.12, 2.9).*

Proof. Clearly the lemma is established if we show that, if for a point z_0 we have $\kappa(z_0, T, \mathfrak{R}) \geqq k$, then there exists an indicator system (z_0, T) of order k in $\mathfrak{R}^0$. Since κ is a lower semi-continuous function of T (cf. §2.5), there exists an $\epsilon_0 > 0$ such that, for every continuous transformation T_* defined on $\mathfrak{R}$ and satisfying $\rho(T_*, T; \mathfrak{R}) < \epsilon_0$, it is true that $\kappa(z_0, T_*, \mathfrak{R}) \geqq k$. Set $\tau = \epsilon_0/3$ and $\epsilon = \epsilon_0/6 = \tau/2$. Now there exists a continuous transformation T_ϵ defined on $\mathfrak{R}$ and satisfying $\rho(T_\epsilon, T, \mathfrak{R}) < \epsilon$ and z_0 not $\in \mathfrak{S}(\infty, T_\epsilon, \mathfrak{R})$[33]. Clearly just a finite number of the components of the set $G(\tau, z_0, T_\epsilon, \mathfrak{R})$ (cf. §2.12) contain points of $T_\epsilon^{-1}(z_0)$—denote by $g_1, \cdots, g_l$ those for which $|t_\epsilon(w) - z_0| = \tau$ for $w \in \bar{g}_i - g_i$ and for $i = 1, \cdots, l$; denote by $g_{l+1}, \cdots, g_m$ those for each of which there exists a point $w_i \in \bar{g}_i - g_i$ such that $|t_\epsilon(w_i) - z_0| < \tau$ for $i = l+1, \cdots, m$. Evidently each of the g_i for $i = 1, \cdots, l$ satisfies the assumptions of Lemma 1 in §2.13, and each of g_i for $i = l+1, \cdots, m$ satisfies the assumptions of Lemma 2 in §2.13. Thus there exist m simply connected Jordan regions $\mathfrak{R}_i$ and m continuous transformations T_i: $z = t_i(w)$ defined on $\mathfrak{R}_i$ and satisfying the conditions set forth in these lemmas. Since the $\bar{g}_i$ are nonoverlapping, so are the $\mathfrak{R}_i$ for $i = 1, \cdots, m$. Set $\Omega = \mathfrak{R} - \sum_{i=1}^m \mathfrak{R}_i$. Define

$$T_\sharp: \quad z = t_\sharp(w) = \begin{cases} t_\epsilon(w) \text{ if } w \in \Omega; \\ t_i(w) \text{ if } w \in \mathfrak{R}_i, \quad i = 1, \cdots, m. \end{cases}$$

Evidently $T_\sharp$ is a continuous transformation defined on $\mathfrak{R}$ and satisfying the

[33] See [30].

following conditions: (i) $\rho(T_\sharp, T; \Omega) < \epsilon$; (ii) $\rho(T_\sharp, T: \mathfrak{R}) < \epsilon_0$; (iii) $N(z_0, T_\sharp, \mathfrak{R}^0) = l$. In view of (ii), (iii) and §2.10, it follows that $k \leqq \kappa(z_0, T_\sharp, \mathfrak{R}) = N(z_0, T_\sharp, \mathcal{N}(T_\sharp, \mathfrak{R}^0)) \leqq l$. Let δ be a positive number such that $O(\delta, t_\sharp, \mathfrak{R}) < \epsilon$ (cf. §2.3). Obviously one may choose Jordan regions $\mathfrak{R}_{\sharp i}$ (not necessarily simply connected) such that $\mathfrak{R}_i \subset \mathfrak{R}_{\sharp i} \subset g_i$ and $\mathfrak{R}_{\sharp i} - \mathfrak{R}_{\sharp i}^0$ lies in a δ-neighborhood of $\bar{g}_i - g_i$ for $i = 1, \cdots, l$. Since each of the $\mathfrak{R}_{\sharp i}$ clearly contains exactly one point $w_{\sharp i} \in T_\sharp^{-1}(z_0)$ in its interior and none on its boundary (cf. §2.13), it follows that $\mu(z_0, T_\sharp, \mathfrak{R}_{\sharp i}) = j(w_{\sharp i}, T_\sharp)$ for $i = 1, \cdots, l$ (cf. §2.10). On the other hand, it is clear from the choice of the $\mathfrak{R}_{\sharp i}$ and from (i) above that $\epsilon(z_0, T_\sharp, \mathfrak{R}_{\sharp i}) > \epsilon > \rho(T, T_\sharp; \mathfrak{R}_{\sharp i} - \mathfrak{R}_{\sharp i}^0)$ for $i = 1, \cdots, l$ (cf. §2.7). Thus, by Lemma 2 in §2.7 it follows that $\mu(z_0, T, \mathfrak{R}_{\sharp i}) = \mu(z_0, T_\sharp, \mathfrak{R}_{\sharp i})$ for $i = 1, \cdots, l$. Since clearly $T_\sharp^{-1}(z_0) \cdot \mathfrak{R}^0 = \sum_{i=1}^{l} w_{\sharp i}$, and $\kappa(z_0, T_\sharp, \mathfrak{R}) \geqq k$, it follows that at least k of the points $w_{\sharp i}$ are in the set $\mathcal{N}(T_\sharp, \mathfrak{R}^0)$. Thus, for some k of the Jordan regions $\mathfrak{R}_{\sharp i}$, it is true that $\mu(z_0, T, \mathfrak{R}_{\sharp i}) \neq 0$; these k mutually exclusive Jordan regions constitute an indicator system (z_0, T) of order k in $\mathfrak{R}^0$. So the lemma is proven.

2.15. Lemma. *If $\mathfrak{R}_*$ be any indicator region (z_0, T) in $\mathfrak{R}^0$, then there exists an indicator region $\mathfrak{R}_\sharp \subset R_*^0$ such that $M(t(w) - z_0, \mathfrak{R}_\sharp) \leqq \frac{1}{2} M(t(w) - z_0, \mathfrak{R}_*)$ (cf.* §2.3).

Proof. Set $\tau = \frac{1}{2}\epsilon(z_0, T, \mathfrak{R}_*)$ (cf. §2.7); clearly $\tau > 0$. Consider the set $G_* = G(\tau, z_0, T, \mathfrak{R} \cdot \mathfrak{R}_*^0)$ (cf. §2.12). Since it is clear that $G_* \in \mathfrak{G}_*^0$, it follows that but a finite number of the components of G_* contain points of $T^{-1}(z_0)$—denote these components by $g_1, \cdots, g_k$. Since $|t(w) - z_0| = \tau > 0$ for $w \in \bar{g}_i - g_i$ and for $i = 1, \cdots, k$ (cf. §2.12), it follows that one may choose in each g_i a Jordan region $\mathfrak{R}_i$ containing the set $T^{-1}(z_0) \cdot g_i$ in its interior. By Lemma 5 in §2.7 it follows that $\mu(z_0, T, \mathfrak{R}^*) = \sum_{i=1}^{k} \mu(z_0, T, \mathfrak{R}_i)$. Now since $\mu(z_0, T, \mathfrak{R}_*) \neq 0$ (cf. §1.15), it follows that for some $\mathfrak{R}_i$—denote it by $\mathfrak{R}_\sharp$—$\mu(z_0, T, \mathfrak{R}_\sharp) \neq 0$. Clearly $\mathfrak{R}_\sharp$ is an indicator region (z_0, T) with the required property.

2.16. Lemma. *Every indicator region (z, T) in $\mathfrak{R}^0$ contains an essential maximal model continuum of z under T in its interior; hence $\phi(z, T, \mathfrak{R}^0) \geqq \psi(z, T, \mathfrak{R}^0)$ (cf.* §2.9).

Proof. If $\mathfrak{R}_0$ be any indicator region (z_0, T), the lemma in §2.15 insures the existence of a sequence $\mathfrak{R}_1, \cdots, \mathfrak{R}_n, \cdots$ of indicator regions (z_0, T) such that $\mathfrak{R}_{n+1} \subset \mathfrak{R}_n^0$ for $n = 0, 1, 2, \cdots$, and $\lim M(t(w) - z_0, \mathfrak{R}_n) = 0$. The reader will verify without trouble that the set $\prod_{n=0}^{\infty} \mathfrak{R}_n$ is an essential maximal model continuum for z_0 under T in $\mathfrak{R}_0^0$.

2.17. In view of the results in §§2.9, 2.14, 2.16 the theorem stated in §2.6 is now proven. Moreover, from the definitions of $\kappa(z, T, \mathfrak{R})$ (cf. §1.12) and of $\kappa(z, T, \mathfrak{R}^0)$ (cf. §1.13), we have, using the theorem in §2.6, the following

Compatibility theorem. $\kappa(z, T, \mathfrak{R}) = \kappa(z, T, \mathfrak{R}^0)$.

2.18. In the sequel we shall assume that T is a bounded continuous transformation defined on a bounded domain $\mathfrak{D}$ in the w-plane (cf. §1.13). Then $\kappa(z, T, \mathfrak{D})$ (cf. §1.13) possesses the following strong lower semi-continuity properties.

THEOREM. *Given a sequence of domains $\mathfrak{D}_n$ which will fill up $\mathfrak{D}$ from the interior*[34]*, and a sequence of continuous transformations T_n defined on $\mathfrak{D}_n$ and such that, for every closed set F in $\mathfrak{D}$ it is true that* $\lim \rho(T_n, T; F)=0$. *Then, for every z,* $\liminf \kappa(z, T_n, \mathfrak{D}_n) \geqq \kappa(z, T, \mathfrak{D})$.

Proof. Clearly it is sufficient to show that, if $\kappa(z, T, \mathfrak{D}) \geqq k$, then $\kappa(z, T_n, \mathfrak{D}_n) \geqq k$ for all n sufficiently large. If $\kappa(z, T, \mathfrak{D}) \geqq k$, it follows by the theorem in §1.17 and by the definition of essential maximal model continua for z under T (cf. §1.16), that there exists in $\mathfrak{D}$ an indicator system $\mathfrak{S}$ of order k for z under T. Since the $\mathfrak{D}_n$ fill up $\mathfrak{D}$ from the interior, and since the Jordan regions $\mathfrak{R} \in \mathfrak{S}$ are closed, it follows that $\mathfrak{S}$ is an indicator system (z, T) of order k in $\mathfrak{D}_n$ for all n sufficiently large. But, from the lemma in §2.8 and the assumption that $\lim \rho(T_n, T; \mathfrak{R})=0$ for every $\mathfrak{R} \in \mathfrak{S}$, it follows that $\mathfrak{S}$ is also an indicator system (z, T_n) of order k in $\mathfrak{D}_n$ for all n sufficiently large. Consequently, by the lemma in §2.16 and the theorem in §1.17 it follows that $\kappa(z, T_n, \mathfrak{D}_n) \geqq k$ for all n sufficiently large, and the theorem is established.

2.19. We turn, next, to the task of verifying the results stated in §1.19; indeed, we shall establish more comprehensive results than those which have been stated there. Our results depend upon the

LEMMA. *If $\mathfrak{R}_1$ and $\mathfrak{R}_2$ be any two indicator regions (z, T) in $\mathfrak{D}$ (cf. §1.15) which contain the same essential maximal model continuum $\sigma(z, T)$ (cf. §1.16), and are contained in a neighborhood of $\sigma(z, t)$ which is free of points belonging to other essential maximal model continua of z under T*[35]*, then* $\mu(z, T, \mathfrak{R}_1) = \mu(z, T, \mathfrak{R}_2)$.

Proof. First, assume that $\mathfrak{R}_2 \subset \mathfrak{R}_1^0$. Consider the closed set $\mathfrak{R}_1 - \mathfrak{R}_2^0$; clearly this set consists of a finite number of Jordan regions $\mathfrak{R}_3, \cdots, \mathfrak{R}_k$, such that $\epsilon(z, T, \mathfrak{R}_i)>0$ for $i=3, \cdots, k$, and no $\mathfrak{R}_i$ for $i=3, \cdots, k$ contains any essential maximal model continuum for z under T; thus (cf. §2.16), it follows that $\mu(z, T, \mathfrak{R}_i)=0$ for $i=3, \cdots, k$. But by Lemma 5 in §2.7, it is true that $\mu(z, T, \mathfrak{R}_1)=\sum_{i=2}^{k}\mu(z, T, \mathfrak{R}_i)=\mu(z, T, \mathfrak{R}_2)$. Thus the lemma is established in this special case. In the general case, since $\sigma(z, T)$ is an essential maximal model continuum in the interior of $\mathfrak{R}_1 \cdot \mathfrak{R}_2$, there exists (cf. §1.16) in the interior of $\mathfrak{R}_1 \cdot \mathfrak{R}_2$ an indicator region $\mathfrak{R}_3$ containing, of course, $\sigma(z, T)$. The preceding reasoning shows that $\mu(z, T, \mathfrak{R}_1)=\mu(z, T, \mathfrak{R}_3)=\mu(z, T, \mathfrak{R}_2)$, and the lemma is now established.

---

[34] See [22].

[35] If $\sigma(z, T)$ consists of a single point w, then clearly $w \in \mathfrak{N}(T, \mathfrak{D})$.

2.20. The assertions made in §1.19 now follow as a corollary to the lemma in §2.19. Moreover, we may now introduce a *general local index* as follows: If $\sigma(z, T)$ be any essential maximal model continuum which has a neighborhood free of points belonging to other essential maximal model continua of z under T, then the quantity $\mu(z, T, \mathfrak{R})$ is independent of the choice of an indicator region $\mathfrak{R}$ for z under T in this neighborhood([36]) by the preceding lemma; we denote its value by $j_*(\sigma, T)$. Clearly, if w be any point of $\mathcal{N}(T, \mathcal{D})$ (cf. §1.18)—that is, if w constitutes an essential maximal model continuum σ—then $j_*(\sigma, T) = j(w, T)$ (cf. §1.19). Thus the general local index coincides with the essential local index on the set $\mathcal{N}(T, \mathcal{D})$. It is quite interesting to observe that, while the general local index is topologically and aesthetically more beautiful, it is the essential local index which plays the fundamental rôle in our applications (cf. Chapter IV). This happens, of course, because of metric assumptions (cf. Chapter IV) which reduce the essential maximal model continua which do not consist of single points to insignificance. But one interesting and enlightening fact concerning the general local index is contained in the

THEOREM. *If $\mathfrak{R}$ be any Jordan region in $\mathcal{D}$ containing but a finite number of essential maximal model continua $\sigma_1(z, T), \cdots, \sigma_k(z, T)$ and for which $\epsilon(z, T, \mathfrak{R}) > 0$ (cf. §2.7), then $\mu(z, T, \mathfrak{R}) = \sum_{i=1}^{k} j_*(\sigma_i, T)$.*

Proof. In $\mathfrak{R}^0$ choose k mutually exclusive indicator regions $\mathfrak{R}_i$ for z under T such that $\sigma_i \subset \mathfrak{R}_i$ for $i = 1, \cdots, k$. The closed set $\mathfrak{R} - \sum_{i=1}^{k} \mathfrak{R}_i^0$ consists of a finite number of Jordan regions $\mathfrak{R}_{k+1}, \cdots, \mathfrak{R}_l$, such that $\epsilon(z, T, \mathfrak{R}_i) > 0$ for $i = k+1, \cdots, l$. In view of the lemmas in §2.16 and §2.19 it is evident that

$$\mu(z, T, \mathfrak{R}_i) = \begin{cases} j_*(\sigma_i, T) \text{ for } i = 1, \cdots, k, \\ 0 \text{ for } i = k+1, \cdots, l. \end{cases}$$

But by Lemma 5 in §2.7 it follows that $\mu(z, T, \mathfrak{R}) = \sum_{i=1}^{l} \mu(z, T, \mathfrak{R}_i)$. Thus $\mu(z, T, \mathfrak{R}) = \sum_{i=1}^{k} j_*(\sigma_i, T)$, and the theorem is proven.

2.21. The following corollary to the results in §2.20 will be useful in the applications of our theory (cf. Chapter IV).

THEOREM. *Given a bounded continuous transformation T defined on a Jordan region $\mathfrak{R}$ (cf. §1.8). Let $\overline{E}$ be any set in the z-plane satisfying the conditions: (i) $|\overline{E} \cdot \overline{\mathfrak{K}}(\infty, T, \mathfrak{R})| = 0$; (ii) $|\overline{E} \cdot T(\mathfrak{R} - \mathfrak{R}^0)| = 0$; (iii) $\kappa(z, T, \mathfrak{R}) = N(z, T, \mathcal{E} \cdot \mathfrak{R}^0)$ a.e. on $\overline{E}$. Then $\mu(z, T, \mathfrak{R}) = \nu(z, T, \mathfrak{R}^0)$ a.e. on $\overline{E}$ (cf. §§1.10–1.14, 1.31).*

Proof. In view of the assumptions, for almost every point $z \in \overline{E}$ it is true that $\kappa(z, T, \mathfrak{R}) = N(z, T, \mathcal{E} \cdot \mathfrak{R}^0) < +\infty$, and that $\epsilon(z, T, \mathfrak{R}) > 0$ (cf. §2.7). For such points, we have $\nu(z, T, \mathfrak{R}^0) = \sum j(w, T)$ for $w \in \mathcal{E} \cdot \mathfrak{R}^0 \cdot T^{-1}(z)$ by definition.

([36]) In particular, if $\sigma(z, T)$ consists of a single point $w \in \mathcal{N}(T, \mathcal{D})$ (see §1.18), then we may always choose $\mathfrak{R}$ as a simply connected Jordan region. However, simple examples show that generally $\mathfrak{R}$ may not be chosen as a simply connected Jordan region.

On the other hand, every maximal model continuum which is essential for such a point z (cf. §1.16) consists of a single point which is clearly in $\mathcal{N}$ (cf. §1.18). In view of the facts stated in §2.20, the theorem is now obvious.

2.22. For the applications (cf. Chapter IV) we shall need to know that the sets $\mathcal{E}(T, \mathcal{D})$ and $\mathcal{N}(T, \mathcal{D})$ (cf. §1.18) are Borel sets, and that the function $j(w, T)$ (cf. §1.19) is a Baire function. We proceed to develop point-set identities which will reveal these facts (cf. §§2.23–2.26).

2.23. Let n be any positive integer, m any integer. Denote by $E_{m,n}$ the set of all points $w \in \mathcal{D}$ for each of which there exists a Jordan region $\mathfrak{R} \subset \mathcal{D}$ satisfying the following conditions: (i) $w \in \mathfrak{R}^0$; (ii) $\mathfrak{R}$ is contained in the open disk with center w and radius $1/n$; (iii) $\mu(t(w), T, \mathfrak{R}) = m$. From Lemma 2 in §2.7 it follows that $E_{m,n}$ is an open set. Clearly $E_{m,n} \supset E_{m,n+1}$ for $n = 1, 2, \cdots$. The reader will easily verify that

$$\mathcal{E}(T, \mathcal{D}) = \prod_{n=1}^{\infty} \sum_{m \neq 0} E_{m,n}.$$

2.24. For purposes in the sequel, we need another tool which we now introduce. Given a bounded domain $\mathcal{D}$ in the w-plane, choose an oriented([37]) square Q containing $\mathcal{D}$ in its interior. Let D_{p_j} denote a subdivision of Q into p_j^2 congruent (oriented) squares s, where p_j is the jth positive prime. If, now, w be any point in $\mathcal{D}$, it follows that there exists a $j_0 = j_0(w)$ such that, for every $j > j_0$ the point w is comprised in the interior of some square of D_{p_j} (see Radó [5, p. 197]).

2.25. Given a bounded continuous transformation T defined on a bounded domain $\mathcal{D}$ (cf. §1.13), let U denote the set of all points $w \in \mathcal{D}$ each of which has a neighborhood $\mathfrak{N}(w)$ such that $\mathfrak{N}(w) - w$ contains no essential maximal model continuum for $t(w)$ under T (cf. §1.16). The reader will easily verify that (cf. §§1.11, 2.24)

$$U = \sum_{j=1}^{\infty} \sum_{s \in D_{p_j}, s^0 \subset \mathcal{D}} [T^{-1}(\overline{\mathfrak{K}}(0, T, s)) - T^{-1}(\overline{\mathfrak{K}}(z, T, s))] \cdot s^0.$$

Thus U is a Borel set, since the sets $\overline{\mathfrak{K}}(k, T, s)$ for $k < +\infty$ are open by Lemma 1 in §2.2, and the inverse (cf. §1.8) of an open set under the continuous transformation T is open.

2.26. The reader will now see without difficulty that (cf. §§2.23, 2.25)

$$\mathcal{N}(T, \mathcal{D}) = U \cdot \mathcal{E} = U \cdot \prod_{n=1}^{\infty} \sum_{m \neq 0} E_{m,n}.$$

Further, if $\mathcal{N}_m = \mathcal{N}_m(T, \mathcal{D})$, $m \neq 0$, denotes the set of points $w \in \mathcal{D}$ where $j(w, T) = m$ (cf. §1.19), then the reader will easily show that

([37]) See ([4]).

$$\mathcal{N}_m(T, \mathcal{D}) = U \cdot \prod_{n=1}^{\infty} E_{m,n}, \qquad m \neq 0.$$

In view of the point-set identities in §§2.23, 2.25 and 2.26, the assertions made in §2.22 are true.

2.27. Given a set $\overline{E}$ in the z-plane, we shall denote by $g(z, \overline{E})$ its characteristic function—that is, the function whose value is 1 for $z \in \overline{E}$ and 0 otherwise. Obviously $g(z, \overline{E})$ is a measurable function of z if and only if $\overline{E}$ is a measurable set.

2.28. The reader will verify without difficulty that, if E be any set in the w-plane, then (cf. §§1.10, 2.24)

$$\sum_{s \in D_{p_j}, s^0 \subset \mathcal{D}} g(z, T(s^0 \cdot E)) \xrightarrow[j \to \infty]{} N(z, T, E \cdot \mathcal{D}).$$

Consequently $N(z, T, E \cdot \mathcal{D})$ is a measurable function of z whenever the set $T(s^0 \cdot E)$ is measurable for every choice of the square s such that $s^0 \subset \mathcal{D}$ (cf. §2.27).

2.29. If E be any Borel set in the w-plane, then $s^0 \cdot E$ is a Borel set for every choice of the square s in the w-plane. Consequently the set $T(s^0 \cdot E)$ is measurable for every square s (see Kuratowski [1, p. 249](38)). Since the sets $\mathcal{E}(T, \mathcal{D})$, $\mathcal{N}(T, \mathcal{D})$, $\mathcal{N}_m(T, \mathcal{D})$, $m \neq 0$, are all Borel sets (cf. §§2.22–2.26), we have the

THEOREM. *If B be any Borel set in the w-plane, then the functions $N(z, T, \mathcal{E} \cdot B)$, $N(z, T, \mathcal{N} \cdot B)$, $N(z, T, \mathcal{N}_m \cdot B)$, $m \neq 0$, are all measurable.*

2.30. For the applications (cf. Chapter IV), we want to discuss the measurability of $\nu(z, T, \mathcal{D})$ (cf. §1.31). We have the

LEMMA. $\nu(z, T, \mathcal{D}) = N(z, T, \mathcal{N}_{+1} \cdot \mathcal{D}) - N(z, T, \mathcal{N}_{-1} \cdot \mathcal{D})$ *for* z *not* $\in \overline{\mathfrak{S}}(\infty, T, \mathcal{E} \cdot \mathcal{D}) + T(\mathcal{N}_*)$.

Proof. Our assumptions imply that $\nu(z, T, \mathcal{D}) = \sum j(w, T)$ for $w \in \mathcal{E} \cdot \mathcal{D} \cdot T^{-1}(z)$, and that $|j(w, T)| \leq 1$ for $w \in \mathcal{E} \cdot \mathcal{D} \cdot T^{-1}(z)$ (cf. §§1.16, 1.20). Thus the identity is obvious.

Since $N(z, T, \mathcal{E} \cdot \mathcal{D})$ is a measurable function of z (cf. §2.29), it follows that the set $\overline{\mathfrak{S}}(\infty, T, \mathcal{E} \cdot \mathcal{D})$ is measurable. Now we have $\nu(z, T, \mathcal{D}) = 0$ for $z \in \overline{\mathfrak{S}}(\infty, T, \mathcal{E} \cdot \mathcal{D})$. In §2.39 we shall show that the set $T(\mathcal{N}_*)$ is denumerable. Consequently we have the

THEOREM. *$\nu(z, T, \mathcal{D})$ is a measurable function of z.*

2.31. We have yet to establish the fact that the set $\mathcal{N}_*$ defined in §1.20

(38) Since, in later chapters, we consider only transformations which satisfy certain metric restrictions, we could very well do without this general theorem on the measurability of the continuous image of a Borel set.

is *denumerable*. In order to do this, we require some results concerning single-valued continuous kth roots of a complex function of a complex variable. If $t(w)$ is a single-valued continuous function of w on a connected set $\mathfrak{C}$, then we say that a function $f(w)$ is a single-valued continuous kth root of $t(w)$ on $\mathfrak{C}$ if (i) $f(w)$ is single-valued and continuous on $\mathfrak{C}$, and (ii) $f(w)^k = t(w)$ on $\mathfrak{C}$.

2.32. LEMMA. *Let $\mathfrak{C}$ denote a connected set in the w-plane. If $t(w)$ be any bounded continuous function defined on $\mathfrak{C}$ for which there exists a single-valued continuous kth root $f(w)$ defined on $\mathfrak{C}$, then $O(f, \mathfrak{C}) \leqq 3[O(t, \mathfrak{C})]^{1/k}$ (cf. §2.3)*[39].

Proof. For $k=1$ the lemma is obvious, since then $f(w)=t(w)$, $w \in \mathfrak{C}$ (cf. §2.31); consequently $O(f, \mathfrak{C}) = O(t, \mathfrak{C})$. So assume that $k>1$. It is now convenient to consider two cases. I. First, assume that

$$M(t, \mathfrak{C}) > 2O(t, \mathfrak{C}). \tag{1}$$

Consider the continuous transformation T defined by $z=t(w)$, $w \in \mathfrak{C}$ (cf. §1.8). It is clear that $T(\mathfrak{C})$ is contained in a closed disk $\mathfrak{D}$ of radius $O(t, \mathfrak{C})$ whose center is at a distance exceeding $2O(t, \mathfrak{C})$ from the origin in the z-plane. Hence $z=0$ is exterior to $\mathfrak{D}$ and we have a single-valued continuous kth root of z in $\mathfrak{D}$ which we denote by $g(z)$, $z \in \mathfrak{D}$. Clearly $g(t(w))$, $w \in \mathfrak{C}$, is a single-valued continuous kth root of $t(w)$ on $\mathfrak{C}$. Thus $g(t(w)) = af(w)$, $w \in \mathfrak{C}$, where a is a kth root of unity; hence it is clear that $O(f, \mathfrak{C}) = O(g, T(\mathfrak{C})) \leqq O(g, \mathfrak{D})$. Now $g(z)$ is an analytic function of z in $\mathfrak{D}$ and therefore we have, for any two points z', z'' in $\mathfrak{D}$,

$$g(z'') - g(z') = \int_{z'}^{z''} g'(z)dz, \tag{2}$$

where the path of integration may be chosen as a straight segment joining z' and z''. Evidently

$$g'(z) = \frac{1}{k[g(z)]^{k-1}}. \tag{3}$$

In view of (1), (2), (3) and the fact that $k>1$, we have

$$O(f, \mathfrak{C}) \leqq O(g, \mathfrak{D}) \leqq M(g', \mathfrak{D})2O(t, \mathfrak{C}) \leqq \frac{2}{k}[O(t, \mathfrak{C})]^{1/k} \leqq [O(t, \mathfrak{C})]^{1/k}.$$

Thus the lemma is established in this case. II. Next, assume that $M(t, \mathfrak{C}) \leqq 2O(t, \mathfrak{C})$. Then clearly (cf. §2.3)

$$O(f, \mathfrak{C}) \leqq 2M(f, \mathfrak{C}) \leqq 2[M(t, \mathfrak{C})]^{1/k} \leqq 2^{1+1/k}[O(t, \mathfrak{C})]^{1/k} \leqq 3[O(t, \mathfrak{C})]^{1/k},$$

since $k>1$. The lemma is now established.

---

(39) Simple examples show that the integer 3 may not be replaced by a smaller integer.

2.33. LEMMA. *Let $\mathfrak{R}$ denote any bounded convex region*[40] *in the w-plane. If $t(w)$ be a continuous function defined on $\mathfrak{R}$ for which there exists a single-valued continuous kth root $f(w)$ defined on $\mathfrak{R}$, then $O(\delta, f, \mathfrak{R}) \leqq 3[O(\delta, t, \mathfrak{R})]^{1/k}$ for every positive number δ (cf. §2.3)*[41].

Proof. Given $\delta > 0$, let w' and w'' be any two points of $\mathfrak{R}$ satisfying $|w' - w''| \leqq \delta$. Denote by C the closed line segment with end points w' and w''. Now $C \subset \mathfrak{R}$ since $\mathfrak{R}$ is convex. Thus it follows by the lemma in §2.32 that

$$|f(w') - f(w'')| \leqq O(f, C) \leqq 3[O(t, C)]^{1/k} \leqq 3[O(\delta, t, \mathfrak{R})]^{1/k}.$$

Consequently $O(\delta, f, \mathfrak{R}) \leqq 3[O(\delta, t, \mathfrak{R})]^{1/k}$, as asserted.

2.34. LEMMA. *Let $\mathfrak{R}$ denote any Jordan region in the w-plane. If T be a continuous transformation defined on $\mathfrak{R}$ for which $\mu(0, T, \mathfrak{R}) = \pm k$, $k > 0$, and the set $T^{-1}(0)$ consists of a single point $w_0 \in \mathfrak{R}^0$, then $t(w)$ possesses a single-valued continuous kth root $f(w)$ in $\mathfrak{R}$.*

Proof. Let l denote a closed line segment joining w_0 to the boundary of $\mathfrak{R}$. In the simply connected domain $\mathfrak{R}^0 - l$ it is true that $t(w) \neq 0$; consequently[42] there exists a single-valued continuous argument $\phi(w)$ for $t(w)$ in $\mathfrak{R}^0 - l$. Now

$$t(w) = |t(w)| [\cos \phi(w) + i \sin \phi(w)], \qquad w \in \mathfrak{R}^0 - l.$$

Set

$$g(w) = |t(w)|^{1/k} \left[\cos \frac{\phi(w)}{k} + i \sin \frac{\phi(w)}{k}\right], \qquad w \in \mathfrak{R}^0 - l.$$

Using Lemma 1 in §2.7 and the assumption that $\mu(0, T, \mathfrak{R}) = \pm k$, the reader will easily show that, for every $w_* \in \mathfrak{R}$, $\lim g(w)$ for $w \in \mathfrak{R}^0 - l$, $w \to w_*$ exists. Define

$$f(w_*) = \lim g(w), \qquad \text{for } w \in \mathfrak{R}^0 - l,\ w \to w_* \in \mathfrak{R}.$$

The reader will find that $f(w)$ is a single-valued continuous kth root of $t(w)$ on $\mathfrak{R}$.

2.35. LEMMA. *Let $\mathfrak{R}$ be any bounded convex region in the w-plane. Let there be given continuous transformations*

$$T:\ z = t(w), \qquad T_n:\ z = t_n(w), \qquad w \in \mathfrak{R},$$

which satisfy the following conditions: (i) $\lim \rho(T_n, T; \mathfrak{R}) = 0$; (ii) $T_n^{-1}(0)$ *con-*

[40] A bounded convex region is always simply connected and bounded by a Jordan curve.

[41] Indeed, since every simply connected Jordan region is topologically equivalent to a bounded convex region, this result is also true for any simply connected Jordan region. Consequently the results in §§2.35, 2.36, 2.37 are true for any simply connected Jordan region, but we do not need this fact.

[42] We use here the so-called monodromy theorem (see Kerékjártó [1, p. 175]).

sists of a single point $w_n \in \mathfrak{R}^0$; (iii) $\mu(0, T_n, \mathfrak{R}) = \pm k$, $k > 0$. *Then* $t(w)$ *possesses a single-valued continuous kth root* $f(w)$ *on* $\mathfrak{R}$.

Proof. In view of the assumptions on the T_n it follows by the lemma in §2.34 that each $t_n(w)$ possesses a single-valued continuous kth root $f_n(w)$ on $\mathfrak{R}$. Since $\lim \rho(T_n, T; \mathfrak{R}) = 0$, it follows that the functions $t_n(w)$ are uniformly bounded(43) and equi-continuous(44) on $\mathfrak{R}$. Using the lemma in §2.33 one shows that the functions $f_n(w)$ are also uniformly bounded and equi-continuous on $\mathfrak{R}$. It follows then, by the theorem of Arzelà, that there exists a subsequence of the $f_n(w)$ which converges uniformly to a (continuous) function $f(w)$ on $\mathfrak{R}$. Since $f_n(w)^k = t_n(w)$ for every n, it follows that $f(w)^k = t(w)$, $w \in \mathfrak{R}$. That is, $f(w)$ is a single-valued continuous kth root of $t(w)$ on $\mathfrak{R}$.

2.36. Lemma. *Let T be a continuous transformation defined on a bounded convex region* $\mathfrak{R}$ *in the w-plane. Assume that* $z = 0$ *is a point of the z-plane having exactly one essential maximal model continuum under T in* $\mathfrak{R}$, *and this consists of a single point* $w_0 \in \mathfrak{R}^0$ *for which* $j(w_0, T) = \pm k$, $k > 0$ (*cf.* §1.19). *Then there exists a* $\tau > 0$ *such that* $\kappa(z, T, \mathfrak{R}) \geqq k$ *for every z satisfying* $0 < |z| < \tau$.

Proof. Since $\kappa(0, T, \mathfrak{R}) = 1$ (cf. §2.6), it follows from the definition of κ (cf. §1.12) that there exists a sequence of continuous transformations T_n defined on $\mathfrak{R}$ and satisfying the following conditions: (i) $\lim \rho(T_n, T; \mathfrak{R}) = 0$; (ii) $T_n^{-1}(0)$ consists of a single point $w_n \in \mathfrak{R}^0$; (iii) $\mu(0, T_n, \mathfrak{R}) = \pm k$ (cf. Lemma 2, §2.7). But then, from the lemma in §2.35 it follows that $t(w)$ possesses a single-valued continuous kth root $f(w)$ on $\mathfrak{R}$. Consider the continuous transformation T_* defined by $z = f(w)$, $w \in \mathfrak{R}$. The reader will verify without difficulty that the only essential maximal model continuum of $z = 0$ under T_* consists of the single point w_0 and that $\mu(0, T_*, \mathfrak{R}) = \pm 1$. Thus we have $\kappa(0, T_*, \mathfrak{R}) = 1$ (cf. §2.6), and consequently (cf. §2.5) there exists a $\tau_* > 0$ such that $\kappa(z, T_*, \mathfrak{R}) \geqq 1$ for every z satisfying $|z| < \tau_*$. Set $\tau = \tau_*^k$. We assert that $\kappa(z, T, \mathfrak{R}) \geqq k$ for every z satisfying $0 < |z| < \tau$. For let z_0 be any point such that $0 < |z_0| < \tau$; then z_0 has exactly k distinct kth roots $z_1, \cdots, z_k$. Clearly $|z_i| < \tau_*$ and consequently we have $\kappa(z_i, T_*, \mathfrak{R}) \geqq 1$, for $i = 1, \cdots, k$. Thus (cf. §2.6) each z_i possesses at least one essential maximal model continuum $\sigma_i = \sigma(z_i, T_*)$ in $\mathfrak{R}^0$. Since the z_i are distinct, so are the σ_i for $i = 1, \cdots, k$. Now clearly $\sigma_i \subset T^{-1}(z_0)$ for $i = 1, \cdots, k$. Consider the functions

$$g_i(w) = f(w)^{k-1} + f(w)^{k-2} z_i + \cdots + z_i^{k-1}, \qquad w \in \mathfrak{R},$$

for $i = 1, \cdots, k$. Evidently each $g_i(w)$ is continuous on $\mathfrak{R}$, and $g_i(w) = k z_i^{k-1} \neq 0$ for $w \in \sigma_i$ and for $i = 1, \cdots, k$. Consequently there exists a neighborhood $\mathfrak{R}_i$

(43) The sequence of functions $f_n(w)$ is said to be uniformly bounded on $\mathfrak{R}$ if there exists a finite constant M such that $M(f_n, \mathfrak{R}) < M$ for every n.

(44) The sequence of functions $f_n(w)$ is said to be equi-continuous on $\mathfrak{R}$ if, for every $\epsilon > 0$, there exists a $\delta = \delta(\epsilon) > 0$ such that $O(\delta, f_n, \mathfrak{R}) < \epsilon$ for every n.

of σ_i in $\mathfrak{R}$ such that $g_i(w) \neq 0$ for $w \in \mathfrak{R}_i$. Let $\mathfrak{R}_i$ be any indicator region (z_i, T_*) (cf. §1.15) such that $\sigma_i \subset \mathfrak{R}_i \subset \mathfrak{R}_i$ for $i = 1, \cdots, k$. We have the identity

$$t(w) - z_0 = (f(w) - z_i)g_i(w), \qquad w \in \mathfrak{R}.$$

Thus, by Lemma 1 and Lemma 3 in §2.7 it follows that $\mu(z_0, T, \mathfrak{R}_i) = \mu(z_i, T_*, \mathfrak{R}_i) \neq 0$ for $i = 1, \cdots, k$. Thus each σ_i is also an essential maximal model continuum (z_0, T) in $\mathfrak{R}^0$—that is, $\kappa(z_0, T, \mathfrak{R}) \geqq k$, as asserted.

2.37. Lemma. *Let T be a continuous transformation defined on a bounded convex region $\mathfrak{R}$ in the w-plane. Assume that z_0 is a point of the z-plane having exactly one essential maximal model continuum under T in $\mathfrak{R}$, and this consists of a single point $w_0 \in \mathfrak{R}^0$ for which $j(w_0, T) = \pm k$, $k > 0$. Then there exists a $\tau > 0$ such that $\kappa(z, T, \mathfrak{R}) \geqq k$ for every z satisfying $0 < |z - z_0| < \tau$.*

Proof. Consider the transformation T_0 defined by the relation $z = t_0(w) = t(w) - z_0$, $w \in \mathfrak{R}$, and apply the lemma in §2.36.

2.38. Now let T be a bounded continuous transformation defined on a bounded domain $\mathcal{D}$ in the w-plane (cf. §1.8). Given a positive integer k, define, for every square s such that $s^0 \subset \mathcal{D}$, the set $\overline{V}(k, s)$ to be the set of all points z having the following properties: (i) z possesses exactly one essential maximal model continuum $\sigma(z, T)$ in s; (ii) $\sigma(z, T)$ consists of a single point $w \in s^0$; (iii) $j(w, T) = \pm k$ (cf. §1.19). It follows immediately from the lemma in §2.37 that $\overline{V}(k, s)$, for $k > 1$, is an isolated set, and hence denumerable. Next, let $\overline{V}(k, \mathcal{D})$ denote the set of all points z for each of which there exists a point $w \in \mathcal{N} \cdot T^{-1}(z)$ (cf. §1.18) for which the essential index $j(w, T) = \pm k$. The reader will easily verify the identity (cf. §2.24)

$$\overline{V}(k, \mathcal{D}) = \sum_{j=1}^{\infty} \sum_{s \epsilon D_{p_j}, s^0 \subset \mathcal{D}} \overline{V}(k, s).$$

Thus, if $k > 1$, it follows that the set $\overline{V}(k, \mathcal{D})$ is denumerable. Consequently the set

$$\overline{V} = \sum_{k=2}^{\infty} \overline{V}(k, \mathcal{D})$$

is also denumerable.

2.39. The reader will find it easy to verify the

Lemma. $T(\mathcal{N}_*) = \overline{V}$ (*cf.* §1.20).

Thus the image of $\mathcal{N}_*$ under T is a denumerable set.

2.40. Now, for every point z, the set $\mathcal{N}_* \cdot T^{-1}(z)$ is clearly an isolated set. But obviously we have

$$\mathcal{N}_* = \sum_{z \epsilon \overline{V}} \mathcal{N}_* \cdot T^{-1}(z),$$

since $\mathcal{N}_* \subset T^{-1}(\bar{V})$ by the lemma in §2.39. Since $\bar{V}$ contains at most a denumerable number of points, so does $\mathcal{N}_*$; thus the important fact stated in §1.20 is established.

CHAPTER III. METRIC FOUNDATIONS

3.1. We state first a few facts concerning functions of rectangles([45]), due essentially([46]) to Banach (see Banach [1], Radó [5]). Let $F(R)$ be a function of rectangles defined for all closed oriented rectangles in a given bounded domain $\mathcal{D}$. Then $F(R)$ is said to be of ***bounded variation*** in $\mathcal{D}$ if there exists a finite constant M such that

$$\sum_{j=1}^{m} | F(s_j) | \leqq M$$

for every system of closed squares in $\mathcal{D}$ without common interior points. $F(R)$ is said to be ***absolutely continuous*** in $\mathcal{D}$ if, for every $\epsilon > 0$ there exists an $\eta(\epsilon) > 0$ such that

$$\sum_{j=1}^{m} | F(s_j) | \leqq \epsilon$$

for every system of closed squares in $\mathcal{D}$, without common interior points, which satisfy the condition

$$\sum_{j=1}^{m} | s_j | \leqq \eta(\epsilon).$$

If $F(R)$ is absolutely continuous and bounded on $\mathcal{D}$, then it is also of bounded variation in $\mathcal{D}$ (see Radó [5, p. 193]).

3.2. Take a point w_0 in $\mathcal{D}$ and consider all possible sequences $\{s_j\}$ of closed squares such that

$$w_0 \in s_j \subset \mathcal{D}$$

and $|s_j| \to 0$. The least upper bound of lim sup $F(s_j)/|s_j|$, for all such sequences $\{s_j\}$, is the ***upper derivative*** $\bar{F}'(w_0)$. The ***lower derivative*** is defined in a similar way. These derivatives may be equal to $\pm\infty$. They are always measurable. If the upper and lower derivatives are finite and equal at a point w_0, then their common value is the ***derivative*** $F'(w_0)$ of $F(R)$ at the point w_0.

3.3. Let S be a closed square comprised in $\mathcal{D}$, and let $s_1, \cdots, s_m$ be a sys-

---

([45]) Let us state again (see ([4])) that we consider only rectangles having sides parallel to the respective axes.

([46]) The facts listed in §§3.1–3.5 are not stated in exactly the same form in which the reader finds them in Banach [1], but in a form convenient for our purposes. However, the necessary modifications in the proofs of Banach are quite obvious (see Radó [5, p. 192] for further remarks on this subject).

tem of closed squares without common interior points contained in S. We say that $F(R)$ is of *type A* if the following conditions are satisfied(47): (i) $F(s)\geqq 0$ for every square $s\subset\mathfrak{D}$; (ii) $\sum_{j=1}^{m} F(s_j)\leqq F(S)$ for every system of squares $S, s_1, \cdots, s_m$ as described above.

If $F(R)$ is of type A, then its derivative $F'(w)$ exists a.e. in $\mathfrak{D}$ and is summable in every rectangle $R\subset\mathfrak{D}$. Furthermore, we have, for every rectangle $R\subset\mathfrak{D}$,

$$\iint_R F'(w) \leqq F(R).$$

3.4. Given a function of rectangles $F(R)$ in $\mathfrak{D}$, suppose that $F(R)$ is absolutely continuous in $\mathfrak{D}$ and that the derivative $F'(w)$ exists a.e. in $\mathfrak{D}$. Then we have, for every open set $O\subset\mathfrak{D}$, the relation (cf. §2.24)

$$\sum_{s\epsilon D_{p_j}, s^0\subset O} F(s) \xrightarrow[j\to\infty]{} \iint_O F'(w).$$

3.5. Let us now assume only that $F(R)\geqq 0$, and that the derivative $F'(w)$ exists a.e. in $\mathfrak{D}$. Take any open set $O\subset\mathfrak{D}$ and assume that

$$l = \liminf_{j\to\infty} \sum_{s\epsilon D_{p_j}, s^0\subset O} F(s) < +\infty. \tag{4}$$

Then $F'(w)$ is summable in O, and

$$\iint_O F'(w) \leqq l.$$

To prove this assertion, define the functions $\phi_j(w)$ as follows: if s is a square such that $s^0\subset O$, $s\in D_{p_j}$, then $\phi_j(w)=F(s)/|s|$ on the interior of s. Otherwise $\phi_j(w)=0$. Then clearly

$$\phi_j(w)\to F'(w) \text{ a.e. in } O, \tag{5}$$

and (cf. §2.24)

$$\iint_O \phi_j(w) = \sum_{s\epsilon D_{p_j}, s^0\subset O} F(s). \tag{6}$$

In view of the lemma of Fatou(48), (4), (5), (6) imply our assertion.

3.6. Let us consider now a transformation (cf. §1.13)

$$T:\quad z = t(w), \qquad w\in\mathfrak{D}.$$

(47) This type of rectangle functions is not considered explicitly by Banach, but his methods (see Banach [1]) yield easily the desired results.

(48) The lemma of Fatou may be stated as follows. Let $\Phi_n\geqq 0$, $\Phi\geqq 0$ be measurable functions on a bounded measurable set S. Suppose that (i) Φ_n is summable on S; (ii) $\Phi_n\to\Phi$ a.e. on S; (iii) $\liminf\int_S\Phi_n<+\infty$. Then Φ is also summable on S, and $\int_S\Phi\leqq\liminf\int_S\Phi_n$. See, for instance, Saks [1, p. 29].

Since $t(w)$ is bounded by assumption, the set $T(\mathcal{D})$ is contained in some finite open circular disc $\bar{\Delta}$ in the z-plane. Let there be given, in the w-plane, a base set $\mathcal{B}$ with the measurability properties described in §1.21.

3.7. If E is a set comprised in $\mathcal{D}$, we shall use the symbol[49] $\bar{A}(E, k)$ to denote the set of those points z for which $N(z, E)=k$. We have then the identity (cf. §2.24)

$$(7)\qquad \mathcal{D}-\mathcal{B}\subset T^{-1}(\bar{A}(\mathcal{D}\cdot\mathcal{B},+\infty))+\sum_{j=1}^{\infty}\sum_{s^0\in D_{p_j},s^0\subset\mathcal{D}} s^0\cdot T^{-1}(\bar{A}(s^0\cdot\mathcal{B},0)).$$

If $E_1, \cdots, E_m$ are nonoverlapping sets contained in a set $E\subset\mathcal{D}$, then we have the following relations (cf. §§1.10, 2.27):

$$(8)\qquad \sum_{j=1}^{m} N(z, E_j\cdot\mathcal{B}) \leqq N(z, E\cdot\mathcal{B}),$$

$$(9)\qquad \sum_{j=1}^{m} g(z, T(E_j\cdot\mathcal{B})) \leqq N(z, E\cdot\mathcal{B}),$$

$$(10)\qquad \sum_{j=1}^{m}[N(z, E_j\cdot\mathcal{B})-g(z, T(E_j\cdot\mathcal{B}))] \leqq N(z, E\cdot\mathcal{B})-g(z, T(E\cdot\mathcal{B})).$$

If O is any open set comprised in $\mathcal{D}$, then

$$(11)\qquad \sum_{s^0\in D_{p_j},s^0\subset O} g(z, T(s^0\cdot\mathcal{B})) \xrightarrow[j\to\infty]{} N(z, O\cdot\mathcal{B}),$$

and, *a fortiori*,

$$(12)\qquad \sum_{s^0\in D_{p_j},s^0\subset O} N(z, s^0\cdot\mathcal{B}) \xrightarrow[j\to\infty]{} N(z, O\cdot\mathcal{B}).$$

The verification of these simple facts is left to the reader.

3.8. Since $g(z, T(s^0\cdot\mathcal{B}))$ is, by assumption, measurable for every $s^0\subset\mathcal{D}$, it follows from (11) that $N(z, O\cdot\mathcal{B})$ is measurable for every open set $O\subset\mathcal{D}$. As a consequence, the sets $\bar{A}(O\cdot\mathcal{B}, k)$ (cf. §3.7) are measurable.

3.9. Take now any closed set $C\subset\mathcal{D}$, and put $O=\mathcal{D}-C$. We have then $N(z, \mathcal{D}\cdot\mathcal{B})=N(z, C\cdot\mathcal{B})+N(z, O\cdot\mathcal{B})$. In view of §3.8, it follows that $N(z, C\cdot\mathcal{B})$ is measurable on the complement of the set $\bar{A}(\mathcal{D}\cdot\mathcal{B}, +\infty)$. Hence, if this latter set is of measure zero, then it follows that $N(z, C\cdot\mathcal{B})$ is measurable for every closed set $C\subset\mathcal{D}$.

3.10. Now let E be any measurable set comprised in $\mathcal{D}$. We can write $E=e+\Gamma$, where $e\cdot\Gamma=0$, $|e|=0$, and $\Gamma=\sum_{j=1}^{\infty}C_j$, where $C_1\subset C_2\subset\cdots\subset C_j\subset\cdots$ and every C_j is a closed set. We have then

[49] Since the transformation T is fixed throughout this chapter, we shall delete the argument T from the notations introduced in Chapter I; for example, we write $N(z, E)$ instead of $N(z, T, E)$.

$$N(z, E\cdot\mathcal{B}) = N(z, e\cdot\mathcal{B}) + N(z, \Gamma\cdot\mathcal{B}), N(z, C_j\cdot\mathcal{B}) \xrightarrow[j\to\infty]{} N(z, \Gamma\cdot\mathcal{B}).$$

In view of §3.9, we obtain the

LEMMA. *If* $|T(e\cdot\mathcal{B})| = 0$ *whenever* $|e| = 0$, *and if* $|\overline{A}(\mathcal{D}\cdot\mathcal{B}, +\infty)| = 0$, *then, for every measurable set* $E\subset\mathcal{D}$ *the function* $N(z, E\cdot\mathcal{B})$ *is measurable.*

3.11. Let us assume now that T is B.V. $\mathcal{B}$ in $\mathcal{D}$ (cf. §1.23). We can write (cf. §§1.10, 2.24, 2.27)

$$\psi_j(z) = \sum_{s\epsilon D_{p_j}, s^0\subset\mathcal{D}} g(z, T(s^0\cdot\mathcal{B}))\begin{cases} \leqq N(z, \mathcal{D}\cdot\mathcal{B}), \\ \xrightarrow[j\to\infty]{} N(z, \mathcal{D}\cdot\mathcal{B}). \end{cases}$$

Then (cf. §1.22)

$$\iint \psi_j(z) = \sum_{s\epsilon D_{p_j}, s^0\subset\mathcal{D}} G(s).$$

Since T is B.V. $\mathcal{B}$ in $\mathcal{D}$, the summation on the right side of this equality is less than a fixed finite constant independent of j. Hence, by the lemma of Fatou([50]), $N(z, \mathcal{D}\cdot\mathcal{B})$ is summable.

Conversely, suppose that $N(z, \mathcal{D}\cdot\mathcal{B})$ is summable. Then, for any system $s_1, \cdots, s_m$ of closed squares without common interior points and such that $s_1^0, \cdots, s_m^0$ are contained in $\mathcal{D}$, by (9) $\sum_{j=1}^m g(z, T(s_j^0\cdot\mathcal{B}))\leqq N(z, \mathcal{D}\cdot\mathcal{B})$, and hence, by integration

$$\sum_{j=1}^m G(s_j) \leqq \iint N(z, \mathcal{D}\cdot\mathcal{B}).$$

Thus T is B.V. $\mathcal{B}$ in $\mathcal{D}$.

3.12. If T is B.V. $\mathcal{B}$ in $\mathcal{D}$, then by §3.11 the function $N(z, \mathcal{D}\cdot\mathcal{B})$ is summable, and hence the set $\overline{A}(\mathcal{D}\cdot\mathcal{B}, \infty)$ is of measure zero (cf. §3.7).

3.13. If T is B.V. $\mathcal{B}$ in $\mathcal{D}$, then $N(z, \mathcal{D}\cdot\mathcal{B})$ is summable (cf. §3.11), and hence, *a fortiori*, $N(z, R^0\cdot\mathcal{B})$ is summable for every closed rectangle R such that $R^0\subset\mathcal{D}$. We can consider therefore the rectangle function

$$G^*(R) = \iint N(z, R^0\cdot\mathcal{B}).$$

It follows then by integration from (8) that $G^*(R)$ is of type A (cf. §3.3). Hence the derivative of $G^*(R)$ exists a.e. in $\mathcal{D}$. We shall denote this derivative by $D^*(w)$.

It follows by integration from (10) that the rectangle function $G^*(R) - G(R)$ is also of type A, and hence that its derivative exists a.e. in $\mathcal{D}$. Con-

([50]) See ([48]).

sequently the rectangle function $G(R)=G^*(R)-(G^*(R)-G(R))$ also has a derivative a.e. in $\mathfrak{D}$. We shall denote the derivative of $G(R)$ by $D(w)$.

3.14. Suppose T is B.V. $\mathfrak{B}$ in $\mathfrak{D}$. Take any open set $O\subset\mathfrak{D}$. Since $N(z,\mathfrak{D}\cdot\mathfrak{B})$ is now summable (cf. §3.11), it follows from (8), (9), (11), (12) that

$$\sum_{s\epsilon D_{p_j}, s^0\subset O} G(s) \xrightarrow[j\to\infty]{} \iint N(z, O\cdot\mathfrak{B}), \qquad \sum_{s\epsilon D_{p_j}, s^0\subset O} G^*(s) \xrightarrow[j\to\infty]{} \iint N(z, O\cdot\mathfrak{B}),$$

and hence also

$$\sum_{s\epsilon D_{p_j}, s^0\subset O} (G^*(s) - G(s)) \xrightarrow[j\to\infty]{} 0.$$

By 3.5 it follows that $D(w)$ and $D^*(w)$ are summable on O and

$$\iint_O D(w) \leqq \iint N(z, O\cdot\mathfrak{B}).$$

In particular, it follows, for $O=\mathfrak{D}$, that $D(w)$ is summable on $\mathfrak{D}$. If we apply the result in §3.5 to the rectangle function $G^*(R)-G(R)$, for $O=\mathfrak{D}$, then we obtain the inequality

$$\iint_{\mathfrak{D}} (D^*(w) - D(w)) \leqq 0.$$

Since obviously $D^*(w)-D(w)\geqq 0$, it follows that $D^*(w)=D(w)$ a.e. in $\mathfrak{D}$.

3.15. Lemma. *Suppose T is B.V. $\mathfrak{B}$ in $\mathfrak{D}$. Let $\overline{E}$ be a measurable set in the z-plane, such that $N(z,\mathfrak{D}\cdot\mathfrak{B})=0$ a.e. on $\overline{E}$. Then $D(w)=0$ a.e. on the set $T^{-1}(\overline{E})$.*

Proof. Since $T(\mathfrak{D})$ is a bounded set contained in a finite open disc $\overline{\Delta}$ (cf. §3.6), clearly $T^{-1}(\overline{E})=T^{-1}(\overline{\Delta}\cdot\overline{E})$. Hence we can assume that $\overline{E}$ itself is bounded. Take then a sequence of open sets $\overline{O}_n$ such that $\overline{O}_1\supset\overline{O}_2\supset\cdots\supset\overline{O}_n\supset\cdots\supset\overline{E}$, $|\overline{O}_n|\to|\overline{E}|$. Put $T^{-1}(\overline{O}_n)=O_n$, $\Gamma=\prod O_n$, $T^{-1}(\overline{E})=E$. Then clearly $O_1\supset O_2\supset\cdots\supset O_n\supset\cdots\supset\Gamma\supset E$, and hence it is sufficient to show that $D(w)=0$ a.e. on Γ. Using §3.14 we have

$$\begin{aligned}\iint_\Gamma D(w) &\leqq \iint_{O_n} D(w) \leqq \iint N(z, O_n\cdot\mathfrak{B}) = \iint_{\overline{O}_n} N(z, O_n\cdot\mathfrak{B})\\ &\leqq \iint_{\overline{O}_n} N(z,\mathfrak{D}\cdot\mathfrak{B}) \to \iint_{\overline{E}} N(z,\mathfrak{D}\cdot\mathfrak{B}) = 0.\end{aligned}$$

Thus

$$\iint_\Gamma D(w) = 0.$$

Hence $D(w)=0$ a.e. on Γ, and the lemma is proved.

3.16. The preceding result implies various corollaries.

COROLLARY a. *If T is B.V. $\mathcal{B}$ in $\mathcal{D}$, and if $\overline{E}$ is a set of measure zero in the z-plane, then $D(w)=0$ a.e. on $T^{-1}(\overline{E})$.*

COROLLARY b. *If T is B.V. $\mathcal{B}$ in $\mathcal{D}$, then $D(w)=0$ a.e. on $T^{-1}(\overline{A}(\mathcal{D}\cdot\mathcal{B}), +\infty))$ (cf. §§3.12, 3.7).*

COROLLARY c. *If T is B.V. $\mathcal{B}$ in $\mathcal{D}$, then $D(w)=0$ a.e. on $T^{-1}(\overline{A}(\mathcal{D}\cdot\mathcal{B}), 0))$ (cf. §3.7).*

3.17. Finally, let us observe that if T is B.V. $\mathcal{B}$ in $\mathcal{D}$ then it is clearly B.V. $\mathcal{B}$ in every subdomain of $\mathcal{D}$. Corollary b, Corollary c, and the identity (7) imply therefore the following

THEOREM. *If T is B.V. $\mathcal{B}$ in $\mathcal{D}$, then $D(w)=0$ a.e. on $\mathcal{D}-\mathcal{B}$.*

3.18. Suppose T is B.V. $\mathcal{B}$ in $\mathcal{D}$. Consider, in the z-plane, a finite-valued measurable function $H(z)$. We cannot then assert that $H(t(w))$ is measurable in $\mathcal{D}$, but we shall see presently that $H(t(w))D(w)$ is measurable in $\mathcal{D}$.

Proof. Since $T(\mathcal{D})$ is contained in a finite open disc $\overline{\Delta}$ (cf. §3.6), we can restrict ourselves to consider $H(z)$ in $\overline{\Delta}$. Since $H(z)$ is measurable there, we have a sequence of continuous functions $H_n(z)$ in $\overline{\Delta}$ such that $H_n(z)\to H(z)$ on $\overline{\Delta}-\overline{E}$, where $\overline{E}$ is some set of measure zero. Put $E=T^{-1}(\overline{E})$. The function $H_n(t(w))$ is continuous in $\mathcal{D}$, $D(w)$ is measurable in $\mathcal{D}$, and hence our assertion will be proved if we can show that

$$H_n(t(w))D(w) \to H(t(w))D(w) \text{ a.e. in } \mathcal{D}. \tag{13}$$

Now (13) holds a.e. on $\mathcal{D}-E$, since there $H_n(t(w))\to H(t(w))$ and $D(w)$ exists a.e. in $\mathcal{D}$. On E, we have $D(w)=0$ a.e. by Corollary a in §3.16, and hence (13) holds there also.

3.19. Assume again that T is B.V. $\mathcal{B}$ in $\mathcal{D}$. Suppose $H_1(z)$, $H_2(z)$ are finite-valued measurable functions in the z-plane, such that $H_1(z)=H_2(z)$ except on a set $\overline{E}$ of measure zero. Put $E=T^{-1}(\overline{E})$. Then the equation

$$H_1(t(w))D(w) = H_2(t(w))D(w) \tag{14}$$

holds a.e. on $\mathcal{D}-E$, since there $H_1(t(w))=H_2(t(w))$ and $D(w)$ exists a.e. in $\mathcal{D}$. On E, we have $D(w)=0$ a.e. by Corollary a in §3.16 and hence (14) holds a.e. on E. Thus the equation (14) holds a.e. in $\mathcal{D}$.

3.20. Suppose now that T is A.C. $\mathcal{B}$ in $\mathcal{D}$ (cf. §1.23). By §3.1 it follows that T is also B.V. $\mathcal{B}$ in $\mathcal{D}$. From §§3.13, 3.14 and §3.4 it follows then that

$$\iint_O D(w) = \iint N(z, O\cdot\mathcal{B})$$

for every open set $O\subset\mathcal{D}$, in particular, for $O=\mathcal{D}$.

3.21. Now suppose that T is B.V. $\mathcal{B}$ in $\mathcal{D}$, and that

$$\iint_{\mathfrak{D}} D(w) = \iint N(z, \mathfrak{D}\cdot\mathfrak{B}).$$

We wish to show that T is A.C. $\mathfrak{B}$ in $\mathfrak{D}$.

Proof. Take any rectangle R such that $R^0 \subset \mathfrak{D}$. We have then a sequence $\{R_j\}$ of rectangles such that $R, R_1, \cdots, R_j, \cdots$ have no common interior points, and such that $\sum R_j^0 + R^0 = \mathfrak{D}$, except for a set of measure zero. Using §3.14 we have then

$$\iint_{R^0} D(w) \leqq \iint N(z, R^0\cdot\mathfrak{B}), \qquad \iint_{R_j^0} D(w) \leqq \iint N(z, R_j^0\cdot\mathfrak{B}),$$

and hence

$$\begin{aligned}\iint_{\mathfrak{D}} D(w) &= \iint_{R^0} D(w) + \sum_j \iint_{R_j^0} D(w)\\ &\leqq \iint N(z, R^0\cdot\mathfrak{B}) + \sum_j \iint N(z, R_j^0\cdot\mathfrak{B})\\ &\leqq \iint N(z, \mathfrak{D}\cdot\mathfrak{B}) = \iint_{\mathfrak{D}} D(w).\end{aligned}$$

Hence the sign of equality holds throughout in the preceding relations. In particular (cf. §3.13),

$$\iint_{R^0} D(w) = \iint N(z, R^0\cdot\mathfrak{B}) = G^*(R).$$

This shows that $G^*(R)$ and hence $G(R)$ (cf. §1.22) are absolutely continuous in $\mathfrak{D}$ (cf. §3.1). Thus T is A.C. $\mathfrak{B}$ in $\mathfrak{D}$.

3.22. Lemma. *If T is A.C. $\mathfrak{B}$ in $\mathfrak{D}$, and if E is a set of measure zero in $\mathfrak{D}$, then $T(E\cdot\mathfrak{B})$ is also a set of measure zero.*

Proof. Take a sequence of open sets O_n in $\mathfrak{D}$ such that $O_1 \supset O_2 \supset \cdots \supset O_n \supset \cdots \supset E$, $|O_n| \to 0$. Using §3.20, we have

$$|T(O_n\cdot\mathfrak{B})| \leqq \iint N(z, O_n\cdot\mathfrak{B}) = \iint_{O_n} D(w) \to 0.$$

Hence $|T(O_n\cdot\mathfrak{B})| \to 0$. But $T(E\cdot\mathfrak{B}) \subset T(O_n\cdot\mathfrak{B})$ for every n. Thus $T(E\cdot\mathfrak{B})$ is of measure zero.

3.23. Corollary. *If T is A.C. $\mathfrak{B}$ in $\mathfrak{D}$, and if E is a measurable set in $\mathfrak{D}$, then $N(z, E\cdot\mathfrak{B})$ is measurable.*

This follows from §3.22 in view of §§3.10–3.12 and §3.20.

3.24. We assume now that T is A.C. $\mathcal{B}$ in $\mathcal{D}$ and we proceed to prove the transformation formula of §1.29. The proof is obtained in several steps. Let us first assume that $H(z)$ is the characteristic function of an open set $\overline{O}$ in the z-plane, while the set E coincides with an open set $O^* \subset \mathcal{D}$. If we put $O = T^{-1}(\overline{O})$, then obviously the transformation formula is equivalent to the formula

$$\iint_{O \cdot O^*} D(w) = \iint N(z, O \cdot O^* \cdot \mathcal{B}).$$

By §3.20, this formula is correct, since $O \cdot O^*$ is an open set.

3.25. Suppose next that E is an open set $O^* \subset \mathcal{D}$, while $H(z)$ is the characteristic function of a measurable set $\overline{Z}$ in the z-plane. Since T is a bounded continuous transformation, we can assume that $\overline{Z}$ is a bounded set. The function $H(t(w))$ is then the characteristic function of the set $Z = T^{-1}(\overline{Z})$. Take a sequence of bounded open sets $\overline{O}_n$ in the z-plane such that $\overline{O}_1 \supset \overline{O}_2 \supset \cdots \supset \overline{O}_n \supset \cdots \supset \overline{Z}$, $|\overline{O}_n| \to |\overline{Z}|$. Let $H_n(z)$ be the characteristic function of the set $\overline{O}_n$. Then we clearly have $H_n(z) \to H(z)$ a.e. Since $|H_n(z)| \leqq 1$ and since $N(z, \mathcal{D} \cdot \mathcal{B})$ is summable, it follows that

$$\iint H_n(z) N(z, O^* \cdot \mathcal{B}) \to \iint H(z) N(z, O^* \cdot \mathcal{B}). \tag{15}$$

Put $O_n = T^{-1}(\overline{O}_n)$. Then $H_n(t(w))$ is the characteristic function of O_n. By a reasoning analogous to that used in §3.18, we see that $H_n(z) \to H(z)$ a.e. implies that $H_n(t(w))D(w) \to H(t(w))D(w)$ a.e. in $\mathcal{D}$. Since $D(w)$ is summable in $\mathcal{D}$ and since $|H_n(t(w))| \leqq 1$, it follows that

$$\iint_{O^*} H_n(t(w))D(w) \to \iint_{O^*} H(t(w))D(w). \tag{16}$$

By §3.24 we have, for every n,

$$\iint_{O^*} H_n(t(w))D(w) = \iint H_n(z) N(z, O^* \cdot \mathcal{B}). \tag{17}$$

(15), (16), (17) imply the desired result.

3.26. Suppose now that E is an open set $O^* \subset \mathcal{D}$, while $H(z)$ is a bounded measurable function which takes on only a finite number of distinct values. Then $H(z)$ can be written in the form $C_1H_1(z) + C_2H_2(z) + \cdots + C_mH_m(z)$, where $C_1, \cdots, C_m$ are constants, and $H_1(z), \cdots, H_m(z)$ are the characteristic functions of measurable sets. Thus the desired result follows directly from §3.25.

3.27. Let E be an open set $O^* \subset \mathcal{D}$, while $H(z)$ is a bounded measurable function, say $|H(z)| \leqq L$. Then we have a sequence $\{H_n(z)\}$ of functions of the type considered in §3.26, such that $|H_n(z)| \leqq L$ and $H_n(z) \to H(z)$ a.e. in

a finite open disc $\bar{\Delta}$ (cf. §3.6) which contains $T(\mathfrak{D})$. The desired result follows then from §3.26 by well-known theorems on termwise integration.

3.28. Let E be any measurable set in $\mathfrak{D}$, while $H(z)$ is a bounded measurable function. In $\mathfrak{D}$, take a sequence $\{O_n\}$ of open sets such that $O_1 \supset O_2 \supset \cdots \supset O_n \supset \cdots \supset E$, $|O_n| \to |E|$. Clearly then

$$\iint_{O_n} H(t(w))D(w) \to \iint_E H(t(w))D(w). \tag{18}$$

By §3.27 we have

$$\iint_{O_n} H(t(w))D(w) = \iint H(z)N(z, O_n \cdot \mathfrak{B}). \tag{19}$$

Put $\Gamma = \prod O_n$. Then obviously

$$N(z, O_n \cdot \mathfrak{B}) \begin{cases} \leqq N(z, \mathfrak{D} \cdot \mathfrak{B}), \\ \to N(z, \Gamma \cdot \mathfrak{B}) \text{ if } N(z, \mathfrak{D} \cdot \mathfrak{B}) < +\infty. \end{cases}$$

Since the set where $N(z, \mathfrak{D} \cdot \mathfrak{B}) = +\infty$ is of measure zero, and since $H(z)$ is bounded, it follows that

$$\iint H(z)N(z, O_n \cdot \mathfrak{B}) \to \iint H(z)N(z, \Gamma \cdot \mathfrak{B}). \tag{20}$$

But $|\Gamma - E| = 0$, and hence, by §3.22, we have $N(z, (\Gamma - E) \cdot \mathfrak{B}) = 0$ a.e. Consequently

$$N(z, \Gamma \cdot \mathfrak{B}) = N(z, E \cdot \mathfrak{B}) \text{ a.e.} \tag{21}$$

The desired result follows now from (18), (19), (20), (21).

3.29. Let E be a measurable set in $\mathfrak{D}$, while $H(z)$ is a finite-valued, measurable, and non-negative function. For every positive integer n, put

$$H_n(z) = \begin{cases} H(z) & \text{if } H(z) \leqq n, \\ n & \text{if } H(z) > n. \end{cases}$$

By §3.28 we have then

$$\iint_E H_n(t(w))D(w) = \iint H_n(z)N(z, E \cdot \mathfrak{B}).$$

Since all the functions involved are non-negative, the desired result follows by the lemma of Fatou[51].

3.30. Let, finally, E be a measurable set in $\mathfrak{D}$, while $H(z)$ is a finite-valued, measurable function, and such that one of the two integrals

---

[51] See [48].

$$\iint_E H(t(w))D(w), \qquad \iint H(z)N(z, E\cdot \mathfrak{B}) \tag{22}$$

exists. Then one of the integrals

$$\iint_E |H(t(w))|D(w), \qquad \iint |H(z)|N(z, E\cdot \mathfrak{B}) \tag{23}$$

exists also. By §3.29, both of these integrals exist then, and

$$\iint_E |H(t(w))|D(w) = \iint |H(z)|N(z, E\cdot \mathfrak{B}). \tag{24}$$

The existence of the integrals (23) implies, however, the existence of both of the integrals (22). If we put $H^*(z) = |H(z)| - H(z)$, then $H^*(z) \geqq 0$, hence §3.29 applies to $H^*(z)$ and we have

$$\iint_E H^*(t(w))D(w) = \iint H^*(z)N(z, E\cdot \mathfrak{B}), \tag{25}$$

where the existence of these integrals follows from the preceding remarks. The desired result is now obtained by subtracting (25) from (24).

Chapter IV. The generalized Jacobian

4.1. In order to establish the theorems stated in §§1.30–1.36, we need a sequence of lemmas, which we now proceed to consider. The transformation (cf. §1.13)

$$T: \quad z = t(w), \qquad w \in \mathfrak{D},$$

will be fixed until further notice, and hence we shall delete the argument T from the notations introduced in Chapter I([52]). The essential set $\mathcal{E}$, defined in §1.18, will play a fundamental part in the sequel.

4.2. Given a rectangle R such that $R^0 \subset \mathfrak{D}$, we shall denote by $g_1(z, R)$ the characteristic function of the set of those points z where $\nu(z, R^0) \neq 0$ (cf. §1.31). We have then obviously

$$g_1(z, R) \leqq N(z, \mathcal{E}\cdot R^0). \tag{26}$$

Let us take now a subdivision([53]) of R into a finite number of rectangles $R_1, \cdots, R_m$. The reader will easily verify that

$$\sum_{i=1}^{m} [N(z, \mathcal{E}\cdot R_i^0) - g_1(z, R_i)] \leqq N(z, \mathcal{E}\cdot R^0) - g_1(z, R). \tag{27}$$

([52]) See ([49]).

([53]) A subdivision of R consists of a finite number of rectangles $R_1, R_2, \cdots, R_m$ without common interior points such that $R = R_1 + R_2 + \cdots + R_m$.

4.3. Given a rectangle R such that $R^0 \subset \mathcal{D}$, we define $g_2(z, R)$ as the characteristic function of the set $T(\mathcal{E}\cdot R^0)$—that is, of the set where $N(z, \mathcal{E}\cdot R^0) \neq 0$; we define $g_3(z, R)$ as the characteristic function of the set where $\kappa(z, R^0) \neq 0$ (cf. §1.13). We introduce then the rectangle functions[64]

$$G_1(R) = \iint g_1(z, R), \qquad G_2(R) = \iint g_2(z, R) = | T(\mathcal{E}\cdot R^0) |,$$

$$G_3(R) = \iint g_3(z, R).$$

In case $N(z, \mathcal{E}\cdot \mathcal{D})$ is summable, we define

$$G_2^*(R) = \iint N(z, \mathcal{E}\cdot R^0).$$

In case $\kappa(z, \mathcal{D})$ is summable, we define

$$G_3^*(R) = \iint \kappa(z, R^0).$$

The derivatives of these rectangle functions, if they exist, will be denoted by $D_1(w)$, $D_2(w)$, $D_3(w)$, $D_2^*(w)$, $D_3^*(w)$ respectively. We define also a *generalized Jacobian*

$$\mathcal{J}(w) = j(w)D_2(w),$$

where $j(w)$ is the local index defined in §1.19. Since, except for a denumerable set, we have $j(w) = 0, +1, -1$ (cf. §1.20), it follows that, except for a denumerable set, $|\mathcal{J}(w)| \leqq D_2(w)$ whenever $D_2(w)$ exists.

4.4. LEMMA. *If T is B.V. $\mathcal{E}$ in $\mathcal{D}$ (cf. §1.23), then* (i) *$D_2(w)$ and $D_2^*(w)$ exist a.e. in $\mathcal{D}$, are summable there, and $D_2(w) = D_2^*(w)$ a.e. in $\mathcal{D}$*; (ii) *$D_1(w)$ exists a.e. in $\mathcal{D}$ and is summable there.*

Proof. The assertion (i) is a direct consequence of §3.14, with $\mathcal{B} = \mathcal{E}$. To prove (ii) we observe that integration of (27) yields

$$\sum_{i=1}^{m} (G_2^*(R_i) - G_1(R_i)) \leqq G_2^*(R) - G_1(R).$$

This means that the rectangle function $G_2^*(r) - G_1(r)$ is of type A (cf. §3.3). Thus $G_1(r)$ is the difference of the functions $G_2^*(r)$ and $G_2^*(r) - G_1(r)$, each of which is of type A (cf. §3.13). Hence, by §3.3, the derivative $D_1(w)$ exists a.e. in $\mathcal{D}$. Integration of (26) yields $G_1(R) \leqq G_2^*(R)$ and hence $D_1(w) \leqq D_2^*(w)$ a.e. in $\mathcal{D}$. Since $D_2^*(w)$ is summable in $\mathcal{D}$, the summability of $D_1(w)$ follows.

(64) The measurability of the functions involved in the following formulas follows from §§2.18, 2.22–2.30.

4.5. Lemma. *If $\kappa(z, \mathcal{D})$ is summable, then the derivatives $D_2(w)$, $D_2^*(w)$, $D_3(w)$, $D_3^*(w)$ exist a.e. in $\mathcal{D}$, are summable in $\mathcal{D}$, and $D_2(w) = D_2^*(w) = D_3(w) = D_3^*(w)$ a.e. in $\mathcal{D}$.*

Proof. Since $N(z, \mathcal{E}\cdot\mathcal{D}) \leqq \kappa(z, \mathcal{D})$ (cf. §§1.13, 1.18), it follows that $N(z, \mathcal{E}\cdot\mathcal{D})$ is summable, and hence, by §3.11, it follows that T is B.V. $\mathcal{E}$ in $\mathcal{D}$. In view of §4.4 it is sufficient to prove that $D_3(w)$, $D_3^*(w)$ exist and are equal to $D_2(w) = D_2^*(w)$ a.e. in $\mathcal{D}$. Now take any rectangle R such that $R^0 \subset \mathcal{D}$, and let $R_1, R_2, \cdots, R_m$ be rectangles forming a subdivision[55] of R. We have then

$$\sum_{i=1}^{m} \kappa(z, R_i^0) \leqq \kappa(z, R^0),$$

by §1.17. Integration yields (cf. §4.3)

$$\sum_{i=1}^{m} G_3^*(R_i) \leqq G_3^*(R).$$

Thus $G_3^*(R)$ is of type A (cf. §3.3), and hence $D_3^*(w)$ exists a.e. in $\mathcal{D}$. Consider again a rectangle R such that $R^0 \subset \mathcal{D}$, and define (cf. §2.24)

$$\phi_j(z) = \sum_{s \in D_{p_j}, s^0 \subset R^0} \kappa(z, s^0).$$

We have then, by §1.17, $\phi_j(z) \leqq \kappa(z, R^0) \leqq \kappa(z, \mathcal{D})$ and, whenever $\kappa(z, \mathcal{D}) < +\infty$ and hence a.e., $\phi_j(z) \to N(z, \mathcal{E}\cdot R^0)$. These last two relations guarantee termwise integrability of the sequence $\phi_j(z)$ and we obtain

$$\sum_{s \in D_{p_j}, s^0 \subset R^0} G_3^*(s) \xrightarrow[j \to \infty]{} \iint N(z, \mathcal{E}\cdot R^0) = G_2^*(R). \tag{28}$$

We have therefore, by §3.5,

$$\iint_{R^0} D_3^*(w) \leqq G_2^*(R).$$

Since this holds for every rectangle, and thus, in particular, for every square, in $\mathcal{D}$, it follows that $D_3^*(w) \leqq D_2^*(w)$ a.e. in $\mathcal{D}$. Since clearly $G_2^*(R) \leqq G_3^*(R)$ (cf. §4.3), we also have $D_2^*(w) \leqq D_3^*(w)$ a.e. in $\mathcal{D}$. Hence $D_2^*(w) = D_3^*(w)$ a.e. in $\mathcal{D}$. Furthermore, since clearly $G_2(R) \leqq G_3(R) \leqq G_3^*(R)$, the upper and lower derivatives (cf. §3.2) of $G_3(R)$ are comprised a.e. between $D_2(w)$ and $D_3^*(w)$. Since we already know that $D_2(w) = D_2^*(w) = D_3^*(w)$ a.e. in $\mathcal{D}$, it follows that $D_3(w)$ exists a.e. in $\mathcal{D}$ and is equal to $D_2(w)$ a.e. in $\mathcal{D}$. Summing up, we have $D_3^*(w) = D_2^*(w) = D_2(w) = D_3(w)$ a.e. in $\mathcal{D}$, as asserted.

4.6. Lemma. *If T is A.C. $\mathcal{E}$ in $\mathcal{D}$ and if $\kappa(z, \mathcal{D})$ is summable, then $D_1(w) = D_2^*(w)$ a.e. in $\mathcal{D}$.*

(55) See (53).

Proof. Let R be a rectangle such that $R^0 \subset \mathcal{D}$. Let us consider the auxiliary function

$$\alpha_j(z) = \sum_{s \in D_{p_j}, s^0 \subset R^0} g_1(z, s).$$

Clearly $\alpha_j(z) \leqq N(z, \mathcal{E}\cdot R^0) \leqq \kappa(z, \mathcal{D})$ and $\alpha_j(z) \to N(z, \mathcal{E}\cdot R^0)$, provided that $\kappa(z, \mathcal{D}) < +\infty$. Since $\kappa(z, \mathcal{D})$ is summable (and hence $\kappa(z, \mathcal{D}) < +\infty$ a.e.), it follows that the sequence $\alpha_j(z)$ can be integrated termwise. Thus we obtain

$$\sum_{s \in D_{p_j}, s^0 \subset R^0} G_1(s) \xrightarrow[j\to\infty]{} G_2^*(R). \tag{29}$$

The inequality $g_1(z, R) \leqq N(z, \mathcal{E}\cdot R^0)$ yields, by integration,

$$G_1(R) \leqq G_2^*(R). \tag{30}$$

Since T is A.C. $\mathcal{E}$ in $\mathcal{D}$, we have, by §3.20 and §3.14,

$$G_2^*(R) = \iint_R D_2^*(w). \tag{31}$$

Thus the rectangle function $G_1(R)$ has the following properties: (i) $G_1(R)$ is absolutely continuous (by (30) and (31)); (ii) the derivative $D_1(w)$ of $G_1(R)$ exists a.e. in $\mathcal{D}$ (by §4.4). In view of §3.4, these properties (i) and (ii) imply that

$$\sum_{s \in D_{p_j}, s^0 \subset R^0} G_1(s) \xrightarrow[j\to\infty]{} \iint_R D_1(w). \tag{32}$$

(29), (31), (32) imply that

$$\iint_R D_1(w) = \iint_R D_2^*(w).$$

Since R is any rectangle in $\mathcal{D}$, this equality implies that $D_1(w) = D_2^*(w)$ a.e. in $\mathcal{D}$.

4.7. Lemma. *If $\kappa(z, \mathcal{D})$ is summable, then $|\mathcal{J}(w)| = D_2(w)$ a.e. in $\mathcal{D}$ (cf.* §4.3).

Proof. Since $N(z, \mathcal{E}\cdot\mathcal{D}) \leqq \kappa(z, \mathcal{D})$, it follows first, by §3.11, that T is B.V. $\mathcal{E}$ in $\mathcal{D}$, and hence $D_2(w)$ and $\mathcal{J}(w) = j(w)D_2(w)$ exist a.e. in $\mathcal{D}$ (cf. §§3.13, 4.3). Let now w_0 be a point such that $D_2(w_0)$ and $\mathcal{J}(w_0)$ exist and $D_2(w_0) \neq |\mathcal{J}(w_0)|$. Then surely $D_2(w_0) \neq 0$ and $j(w_0) \neq \pm 1$. Except for points which belong to a set of measure zero, these conditions imply that $w_0 \in \mathcal{E}$ (since $D_2(w) = 0$ a.e. on $\mathcal{D} - \mathcal{E}$ by §3.17), and that w_0 is not in $\mathcal{N}$ (since $j(w) = \pm 1$ on $\mathcal{N}$ except for a denumerable set, by §1.20). Hence $w_0 \in \mathcal{E} - \mathcal{N}$. But $T(\mathcal{E} - \mathcal{N})$ is a subset of the set of those points z where $\kappa(z, \mathcal{D}) = +\infty$ (cf. §1.18). This latter set being

of measure zero (since $\kappa(z, \mathcal{D})$ is summable), it follows by §3.16 that $D_2(w)=0$ a.e. on $\mathcal{E}-\mathcal{N}$. Since $D_2(w_0)\neq 0$, the lemma follows.

4.8. We assume now that $T\in K_1(\mathcal{D})$ (cf. §1.32). Applying the theorem of §1.29 with $\mathcal{B}=\mathcal{E}$, we have the fundamental formula

$$\iint_E H(t(w))D_2(w) = \iint H(z)N(z, \mathcal{E}\cdot E)$$

(where E is any measurable set in $\mathcal{D}$ and $H(z)$ is any finite-valued measurable function) provided only that the integrals involved exist. We also know that the existence of one of the integrals involved implies the existence of the other (cf. §1.29).

4.9. Assuming again that $T\in K_1(\mathcal{D})$, let us consider a finite-valued measurable function $H(z)$ such that $H(t(w))\mathcal{J}(w)$ is summable in $\mathcal{D}$. Since $\mathcal{J}(w)=0$ a.e. on $\mathcal{D}-\mathcal{N}$ (cf. §§1.18, 1.19, 3.16, 3.17), this means summability on the set $\mathcal{N}=\mathcal{N}_{+1}+\mathcal{N}_{-1}+\mathcal{N}_*$ (cf. §§1.20, 2.26). But $\mathcal{N}_*$ is denumerable (cf. §1.20). Hence, using §4.8,

$$\begin{aligned}\iint_{\mathcal{D}} H(t(w))\mathcal{J}(w) &= \iint_{\mathcal{N}_{+1}} H(t(w))\mathcal{J}(w) + \iint_{\mathcal{N}_{-1}} H(t(w))\mathcal{J}(w)\\ &= \iint_{\mathcal{N}_{+1}} H(t(w))D_2(w) - \iint_{\mathcal{N}_{-1}} H(t(w))D_2(w)\\ &= \iint H(z)N(z, \mathcal{N}_{+1}) - \iint H(z)N(z, \mathcal{N}_{-1})\\ &= \iint H(z)\left[N(z, \mathcal{N}_{+1}) - N(z, \mathcal{N}_{-1})\right].\end{aligned}$$

By §2.30, $N(z, \mathcal{N}_{+1})-N(z, \mathcal{N}_{-1})=\nu(z, \mathcal{D})$ a.e., and the theorem of §1.32 is proved. The theorem in §1.34 now follows directly by §2.21.

4.10. Concerning the class $K_2(\mathcal{D})$, defined in §1.34, inspection of the lemmas in §§4.5, 4.6, 4.7 yields the

THEOREM. *If $T\in K_2(\mathcal{D})$, then*

$$D_1(w) = D_2(w) = D_3(w) = D_2^*(w) = D_3^*(w) = |\mathcal{J}(w)|$$

a.e. in $\mathcal{D}$.

4.11. We proceed now to prove the closure theorem (cf. §1.35) for the class $K_2(\mathcal{D})$. Using the assumptions and notations of §1.35 let us take a Jordan region $\mathfrak{R}\subset\mathcal{D}$. By assumption[56]

[56] Of course, $t_n(w)$ will be defined in $\mathfrak{R}$ only if n is sufficiently large. In the course of the proof, the subscript n refers to the transformation T_n. For example, $\mathcal{E}_n=\mathcal{E}(T_n, \mathcal{D}_n)$.

$$t_n(w) \rightrightarrows t(w) \text{ on } \mathfrak{R}, \quad \iint_{\mathfrak{R}} | \mathcal{J}(w) - \mathcal{J}_n(w) | \to 0,$$

and hence, *a fortiori*,

$$(33) \quad \begin{gathered} t_n(w) \rightrightarrows t(w) \text{ on } \mathfrak{R}^0, \\ \iint_{\mathfrak{R}^0} | \mathcal{J}(w) - \mathcal{J}_n(w) | \to 0. \end{gathered}$$

Thus (cf. §2.18) a.e.

$$(34) \quad N(z, T, \mathcal{E}\cdot\mathfrak{R}^0) \leqq \kappa(z, T, \mathfrak{R}^0) \leqq \liminf \kappa(z, T_n, \mathfrak{R}^0) = \liminf N(z, T_n, \mathcal{E}_n\cdot\mathfrak{R}^0).$$

Using (33), §4.7 and §4.8, we have

$$(35) \quad \iint N(z, T_n, \mathcal{E}_n\cdot\mathfrak{R}^0) = \iint_{\mathfrak{R}^0} | \mathcal{J}_n(w) | \to \iint_{\mathfrak{R}^0} | \mathcal{J}(w) | \leqq \iint_{\mathcal{D}} | \mathcal{J}(w) |.$$

By the lemma of Fatou([57]), (34) and (35) imply that $\kappa(z, T, \mathfrak{R}^0)$ is summable, and

$$\iint \kappa(z, T, \mathfrak{R}^0) \leqq \iint_{\mathcal{D}} | \mathcal{J}(w) |.$$

If we apply this inequality to a sequence of Jordan regions which fill up $\mathcal{D}$ from the interior (cf. §1.13), it follows, by §2.18, and by the lemma of Fatou, that $\kappa(z, T, \mathcal{D})$ is summable, and

$$(36) \quad \iint \kappa(z, T, \mathcal{D}) \leqq \iint_{\mathcal{D}} | \mathcal{J}(w) |.$$

Since $\kappa(z, T, \mathcal{D})$ is summable, it follows that T is B.V. $\mathcal{E}$ in $\mathcal{D}$ and by §§3.14, 4.3 and (36) we can write

$$\begin{aligned} \iint_{\mathcal{D}} | \mathcal{J}(w) | &\leqq \iint_{\mathcal{D}} D_2(w) \leqq \iint N(z, T, \mathcal{E}\cdot\mathcal{D}) \\ &\leqq \iint \kappa(z, T, \mathcal{D}) \leqq \iint_{\mathcal{D}} | \mathcal{J}(w) |. \end{aligned}$$

Hence the sign of equality holds throughout, and in view of §3.21 the theorem follows.

4.12. We shall need the following well-known theorems in Lebesgue theory. Let $\overline{E}$ be a bounded measurable set in the z-plane. Let $f_n(z)$, $f(z)$ be summable functions on $\overline{E}$, such that $f_n(z) \to f(z)$ a.e. in $\overline{E}$. Let us say that the

---

([57]) See ([48]).

sequence $\{f_n(z)\}$ satisfies the condition U on $\bar{E}$ if, for every $\epsilon>0$, there exists a $\delta=\delta(\epsilon)>0$ such that

$$\iint_{\bar{e}} |f_n(z)| \leqq \epsilon$$

for every measurable subset $\bar{e}$ of $\bar{E}$ such that $|\bar{e}|<\delta$ (see de la Vallée Poussin [1]). Then the following statements are true: (i) we have

$$\iint_{\bar{S}} f_n(z) \to \iint_{\bar{S}} f(z)$$

for every measurable subset $\bar{S}$ of $\bar{E}$ if and only if the sequence $\{f_n(z)\}$ satisfies the condition U on $\bar{E}$; (ii) we have

$$\iint_{\bar{E}} |f_n(z)| \to \iint_{\bar{E}} |f(z)|$$

if and only if the sequence $\{f_n(z)\}$ satisfies the condition U on $\bar{E}$.

4.13. As an immediate corollary to the preceding statements, it follows that

$$\iint_{\bar{E}} |f_n(z) - f(z)| \to 0$$

if and only if

$$\iint_{\bar{E}} |f_n(z)| \to \iint_{\bar{E}} |f(z)|.$$

4.14. We proceed now to prove the theorem of §1.36. Using the notations and assumptions of that theorem, we first note that $\mu(z, T_n, \mathfrak{R})\to\mu(z, T, \mathfrak{R})$ a.e. in the z-plane. Indeed, this holds if z is not in $T(\mathfrak{R}-\mathfrak{R}^0)$ (cf. §2.7), and we have $|T(\mathfrak{R}-\mathfrak{R}^0)|=0$ by assumption. We note also that all the functions $\mu(z, T_n, \mathfrak{R})$, $\mu(z, T, \mathfrak{R})$ are equal to zero outside of a conveniently chosen circular disc (cf. §3.6). In view of §4.13 it is therefore sufficient to prove that

$$\iint |\mu(z, T_n, \mathfrak{R})| \to \iint |\mu(z, T, \mathfrak{R})|.$$

Let E denote the set of those points $w\in\mathfrak{R}^0$ for which $t(w)$ is a point of $T(\mathfrak{R}-\mathfrak{R}^0)$. Since $|T(\mathfrak{R}-\mathfrak{R}^0)|=0$ by assumption, it follows by §3.16 that $\mathcal{J}(w)=0$ a.e. on E. Choosing $H(z)$ as the function[58] sgn $\mu(z, T, \mathfrak{R})$, and $H_n(z)$ as the function sgn $\mu(z, T_n, \mathfrak{R})$, we have, for any fixed z not in $T(\mathfrak{R}-\mathfrak{R}^0)$, the relation $H_n(z)=H(z)$ for large values of n (cf. §2.7). Thus $H_n(z)\to H(z)$ a.e. in the z-plane. But a much stronger relation holds. Indeed, let z_0 be any point

[58] If a is a real number, then sgn $a=+1, -1, 0$ according as $a>0$, $a<0$, $a=0$.

not in $T(\mathfrak{R}-\mathfrak{R}^0)$. Let $\bar{\gamma}$ denote a small closed circular disc which contains z_0 but contains no point of $T(\mathfrak{R}-\mathfrak{R}^0)$. Clearly we have an n_0 such that $H_n(z)=H(z)$ for $z\in\bar{\gamma}$, $n>n_0$. From this we infer that, if z is not in $T(\mathfrak{R}-\mathfrak{R}^0)$ and if $z_n\to z$, then $H_n(z_n)\to H(z)$, and hence

$$H_n(t_n(w)) \to H(t(w)), \text{ if } w \text{ not } \in E. \tag{37}$$

Set

$$I_n' = \iint_{\mathfrak{R}^0} H_n(t_n(w))(\mathcal{J}_n(w) - \mathcal{J}(w)),$$

$$I_n'' = \iint_{\mathfrak{R}^0-E} (H_n(t_n(w)) - H(t(w)))\mathcal{J}(w),$$

$$I_n''' = \iint_{E} (H_n(t_n(w)) - H(t(w)))\mathcal{J}(w).$$

Then clearly

$$\iint_{\mathfrak{R}^0} H_n(t_n(w))\mathcal{J}_n(w) = I_n' + I_n'' + I_n''' + \iint_{\mathfrak{R}^0} H(t(w))\mathcal{J}(w).$$

We have, since $|H_n|\leqq 1$,

$$|I_n'| \leqq \iint_{\mathfrak{R}^0} |\mathcal{J}_n(w) - \mathcal{J}(w)|,$$

hence $I_n'\to 0$ as a direct consequence of our assumptions. By (37) we have, since $|H_n-H|\leqq 2$ and $\mathcal{J}(w)$ is summable, $I_n''\to 0$. Since, by a preceding remark, $\mathcal{J}(w)=0$ a.e. on E, we have $I_n'''=0$. Thus

$$\iint_{\mathfrak{R}^0} H_n(t_n(w))\mathcal{J}_n(w) \to \iint_{\mathfrak{R}^0} H(t(w))\mathcal{J}(w).$$

By the transformation formula of §1.34 this relation yields directly

$$\iint H_n(z)\mu(z, T_n, \mathfrak{R}) \to \iint H(z)\mu(z, T, \mathfrak{R}).$$

This completes the proof, since by the definition of $H_n(z)$ and of $H(z)$ we have

$$H_n(z)\mu(z, T_n, \mathfrak{R}) = |\mu(z, T_n, \mathfrak{R})|, \qquad H(z)\mu(z, T, \mathfrak{R}) = |\mu(z, T, \mathfrak{R})|.$$

Chapter V. The ordinary Jacobian

5.1. In this chapter, we shall consider a bounded continuous transformation T defined on a bounded domain $\mathcal{D}$ by relations of the form (cf. §1.8)

$$T:\quad x = x(u, v), \qquad y = y(u, v), \qquad (u, v) \in \mathcal{D}.$$

We assume that the partial derivatives x_u, x_v, y_u, y_v exist a.e. in $\mathcal{D}$, and we denote the ordinary Jacobian by

$$J(w) = J(u, v) = x_u y_v - x_v y_u.$$

5.2. Let $w_0 = u_0 + iv_0$ (cf. §1.8) be a point of $\mathcal{D}$ and let $\{s_n\}$ be a sequence of closed squares comprised in $\mathcal{D}$. We shall say that the point w_0 and the sequence $\{s_n\}$ jointly satisfy the condition $C(\{s_n\}, w_0)$ if the following facts

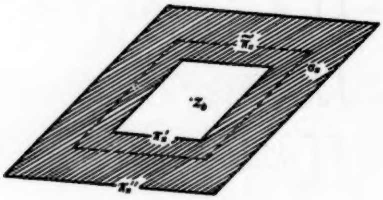


FIG. 1

hold: (i) each s_n is an oriented square with center at w_0; (ii) $\lim |s_n| = 0$; (iii) $J(w_0)$ exists; (iv) if b_n denotes the boundary of s_n—that is, $b_n = s_n - s_n^0$—and if we put

$$\xi(s_n, w_0) = \max_{(u,v)\epsilon b_n} \frac{|x(u, v) - x(u_0, v_0) - (u - u_0)x_u(u_0, v_0) - (v - v_0)x_v(u_0, v_0)|}{[(u - u_0)^2 + (v - v_0)^2]^{1/2}},$$

$$\eta(s_n, w_0) = \max_{(u,v)\epsilon b_n} \frac{|y(u, v) - y(u_0, v_0) - (u - u_0)y_u(u_0, v_0) - (v - v_0)y_v(u_0, v_0)|}{[(u - u_0)^2 + (v - v_0)^2]^{1/2}},$$

then $\lim \xi(s_n, w_0) = 0$, $\lim \eta(s_n, w_0) = 0$.

5.3. Our assumption that x_u, x_v, y_u, y_v exist a.e. in $\mathcal{D}$ (cf. §5.1) implies that, for almost every point $w \in \mathcal{D}$ there exists a sequence $\{s_n\}$ of squares comprised in $\mathcal{D}$ for which the condition $C(\{s_n\}, w)$ is satisfied (see Radó [5, p. 219]).

5.4. Consider now a point w_0 and a sequence $\{s_n\}$ satisfying the condition $C(\{s_n\}, w_0)$. Let us introduce the auxiliary affine transformation

$$\tilde{T}: \begin{array}{l} x = x(u_0, v_0) + (u - u_0)x_u(u_0, v_0) + (v - v_0)x_v(u_0, v_0), \\ y = y(u_0, v_0) + (u - u_0)y_u(u_0, v_0) + (v - v_0)y_v(u_0, v_0), \qquad (u, v) \in \mathcal{D}. \end{array}$$

By a wholly elementary reasoning[59], we obtain the following statements concerning the relation between T and $\tilde{T}$.

Let us denote by z_0 the point $T(w_0) = \tilde{T}(w_0)$. I. Suppose, first, that $J(w_0) \neq 0$. For brevity, denote $\tilde{T}(s_n)$ by $\tilde{\pi}_n$. Then $\tilde{\pi}_n$ is a parallelogram with

[59] See, for instance Radó [1] for a full discussion of practically the same situation.

center at z_0. Since $\tilde{T}$ is an affine transformation, we have

$$| \tilde{\pi}_n | = | J(w_0) | \, | s_n |. \tag{38}$$

As a consequence of condition $C(\{s_n\}, w_0)$, there exist two parallelograms π_n', π_n'' each with center at z_0 and each similar to $\tilde{\pi}_n$ with respect to z_0 (see Figure 1), for which the following statements hold. If σ_n denotes the closed strip bounded by the perimeters of π_n' and π_n'', and if b_n denotes the perimeter of s_n then

$$\begin{cases} T(b_n) \subset \sigma_n, \\ \lim | \sigma_n | / | s_n | = 0. \end{cases} \tag{39}$$

Further, we have for large n (cf. §2.7)

$$\mu(z, T, s_n) = \mu(z, \tilde{T}, s_n) = \begin{cases} \operatorname{sgn} J(w_0) \text{ if } z \in \pi_n', \\ 0 \text{ if } z \text{ not } \in \pi_n''. \end{cases} \tag{40}$$

II. Suppose, second, that $J(w_0)=0$. Then $\tilde{T}(s_n)$ reduces to a straight seg-

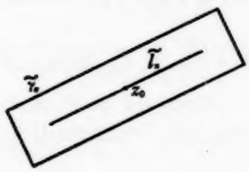


FIG. 2

ment $\tilde{l}_n$ with center at z_0. There exists a closed rectangle $\bar{r}_n$ with center at z_0 (see Figure 2) containing $\tilde{l}_n$ in its interior and such that

$$\begin{cases} T(b_n) \subset \bar{r}_n, \\ \lim | \bar{r}_n | / | s_n | = 0, \end{cases} \tag{41}$$

$$\mu(z, T, s_n) = \mu(z, \tilde{T}, s_n) = 0 \text{ if } z \text{ not } \in \bar{r}_n. \tag{42}$$

5.5. Consider now a continuous transformation $T \in K_3(\mathcal{D})$ (cf. §1.37). For almost every point $w_0 \in \mathcal{D}$ the following statements are true: (i) $D_1(w_0)$, $D_2(w_0)$ both exist, and $D_1(w_0) = D_2(w_0)$ (cf. §4.6); (ii) $J(w_0)$ exists (by assumption); (iii) there exists a sequence $\{s_n\}$ of squares such that the condition $C(\{s_n\}, w_0)$ is satisfied (cf. §5.3); (iv) w_0 is not in the subset of $\mathcal{D} - \mathcal{N}$ where $D_2 > 0$ (cf. §§1.18, 3.17, Corollary b in §3.16). Consider such a point w_0. I. Assume, first, that $J(w_0) \neq 0$. Then we have conditions as described in I of §5.4 and pictured in Figure 1. It follows from (39) and from §2.21 that

$$\mu(z, T, s_n) = \nu(z, T, \overset{0}{s}_n) \text{ for a.e. } z \text{ not } \in \sigma_n. \tag{43}$$

From (40) and (43) it follows (cf. §4.2) that

$$g_1(z, T, s_n) = \begin{cases} 1 & \text{a.e. in } \pi_n', \\ 0 & \text{a.e. not in } \pi_n''. \end{cases}$$

Hence (cf. §4.3) we have $|\pi_n'| \leqq G_1(s_n) \leqq |\pi_n''|$ and consequently

$$(44) \qquad |\bar{\pi}_n| - |\sigma_n| < G_1(s_n) < |\bar{\pi}_n| + |\sigma_n|.$$

Using (38) in §5.4 and conditions (i) and (ii) described in this section, we find from (44) that $D_1(w_0) = |J(w_0)|$, and hence

$$(45) \qquad D_2(w_0) = |J(w_0)|.$$

Thus, in particular, $D_2(w_0) > 0$. In view of condition (iv) of this section, it follows that $w_0 \in \mathcal{N}$. Consequently we have $j(w_0) = \mu(z_0, T, s_n)$ for all n sufficiently large (cf. §1.19). But then, in view of (40) and (45) it follows that

$$(46) \qquad \mathcal{J}(w_0) = J(w_0).$$

II. Assume, second, that $J(w_0) = 0$. Then we have conditions as described in II of §5.4 and as pictured in Figure 2. It follows from (41) and from §2.21 that

$$(47) \qquad \mu(z, T, s_n) = \nu(z, T, \overset{0}{s_n}) \text{ for a.e. } z \text{ not } \in \bar{r}_n.$$

Thus, from (42) and (47) it follows (cf. §4.3) that $G_1(s_n) \leqq |\bar{r}_n|$. Consequently, in view of (41) we have $D_1(w_0) = 0$. Using conditions (i) and (ii) of this section, we find that

$$(48) \qquad \mathcal{J}(w_0) = 0 = J(w_0).$$

In view of (46) and (48) we have $\mathcal{J}(w) = J(w)$ for every point w which satisfies conditions (i)–(iv) described at the beginning of this section. Since the set of points in $\mathcal{D}$ which fail to satisfy these conditions is of measure zero, the theorem in §1.38 is established.

5.6. In order to prove the closure theorem stated in §1.41, we need the following

Lemma. *Assume that the continuous transformation T (cf. §5.1) is such that $\kappa(z, T, \mathcal{D})$ is summable and $J(w)$ exists a.e. in $\mathcal{D}$. Then $D_2(w) \geqq |J(w)|$ a.e. in $\mathcal{D}$.*

Proof. As a consequence of these assumptions, the following conditions hold for almost every point $w_0 \in \mathcal{D}$: (i) $J(w_0)$, $D_2(w_0)$, $D_3^*(w_0)$ exist, and $D_2(w_0) = D_3^*(w_0)$ (cf. §4.5); (ii) there exists a sequence $\{s_n\}$ of squares such that the condition $C(\{s_n\}, w_0)$ is satisfied (cf. §5.3). Let w_0 be such a point. If $J(w_0) = 0$, the lemma is obvious. So assume that $J(w_0) \neq 0$. Then we have conditions as described in I of §5.4 and pictured in Figure 1. In view of (40) we have $\kappa(z, T, s_n^0) \geqq 1$ for z in π_n' (cf. §2.7). Consequently (cf. §4.3), for every n,

$$G_3^*(s_n) = \iint \kappa(z, T, \overset{0}{s_n}) \geqq |\pi_n'| > |\tilde{\pi}_n| - |\sigma_n|.$$

The lemma now follows by condition (i) above, and by (38) and (39) in §5.4.

5.7. We prove presently the closure theorem for the class $K_3(\mathfrak{D})$ stated in §1.41. In view of the closure theorem for the class $K_2(\mathfrak{D})$ stated in §1.35 it is sufficient to prove that $T \in K_3(R^0)$ for every rectangle $R \subset \mathfrak{D}$; for then we can infer from §1.38 that $\mathcal{J}(w) = J(w)$ a.e. in $\mathfrak{D}$, and consequently the closure theorem for the class $K_3(\mathfrak{D})$ follows directly from the closure theorem for the class $K_2(\mathfrak{D})$.

Let R be any rectangle in $\mathfrak{D}$. Since $\mathcal{J}(w, T_n) = J(w, T_n)$ a.e. in $\mathfrak{D}$ (cf. §§1.38, 5.5), we have, from the assumptions in the theorem,

$$\lim \iint_R |J(w, T) - \mathcal{J}(w, T_n)| = 0,$$

and hence

$$\lim \iint_{R^0} |\mathcal{J}(w, T_n)| = \iint_{R^0} |J(w, T)|. \tag{49}$$

Furthermore (cf. §2.18), we have $\kappa(z, T, R^0) \leqq \liminf \kappa(z, T_n, R^0)$. Now $T_n \in K_3(\mathfrak{D}_n)$ and hence[60] $T_n \in K_3(R^0)$ for all n sufficiently large. Thus it follows (cf. §1.37) that

$$\kappa(z, T, R^0) \leqq \liminf N(z, T_n, \mathcal{E}(T_n, \mathfrak{D}_n) \cdot R^0) \text{ for a.e. } z, \tag{50}$$

and, using (49), we have also

$$\iint N(z, T_n, \mathcal{E}(T_n, \mathfrak{D}_n) \cdot R^0) = \iint_{R^0} |\mathcal{J}(w, T_n)| \to \iint_{R^0} |J(w, T)|. \tag{51}$$

By the lemma of Fatou[61], the relations (50) and (51) yield the summability of $\kappa(z, T, R^0)$. Thus, by §5.6, it follows that

$$|J(w, T)| \leqq D_2(w, T) \text{ a.e. in } R^0. \tag{52}$$

Since the summability of $\kappa(z, T, R^0)$ implies the summability of $N(z, T, \mathcal{E} \cdot R^0)$, it follows that T is B.V. $\mathcal{E}$ in $\mathfrak{D}$ (cf. §3.11). Using §3.14, the lemma of Fatou, and the relations (51), (52), we have

$$\iint_{R^0} |J(w, T)| \leqq \iint_{R^0} D_2(w, T) \leqq \iint N(z, T, \mathcal{E} \cdot R^0) \leqq \iint \kappa(z, T, R^0)$$

$$\leqq \liminf \iint N(z, T_n, \mathcal{E}(T_n, \mathfrak{D}_n) \cdot R^0) = \iint_{R^0} |\mathcal{J}(w, T)|.$$

---

(60) Note that $R \subset \mathfrak{D}$ and consequently $R \subset \mathfrak{D}_n$ for all n sufficiently large.

(61) See (48).

Hence the sign of equality holds throughout. In particular,

$$\iint_{R^0} D_2(w, T) = \iint N(z, T, \mathcal{E}\cdot R^0), \tag{53}$$

$$\iint N(z, T, \mathcal{E}\cdot R^0) = \iint \kappa(z, T, R^0). \tag{54}$$

By §3.21, (53) implies that T is A.C. $\mathcal{E}$ in R^0. Since we have $N(z, T, \mathcal{E}\cdot R^0) \leqq \kappa(z, T, R^0)$, the relation (54) implies that $N(z, T, \mathcal{E}\cdot R^0) = \kappa(z, T, R^0)$ a.e. Thus $T \in K_2(R^0)$ (cf. §1.34), but since $J(w, T)$ exists a.e. in $\mathcal{D}$ by assumption, we have $T \in K_3(R^0)$.

5.8. Let us now assume that the continuous transformation T (cf. §5.1) satisfies a Lipschitz condition in $\mathcal{D}$ in the restricted sense described in §1.39. If s be any square comprised in $\mathcal{D}$, then clearly

$$| T(s^0) | \leqq \tfrac{1}{2}\pi L^2 | s |, \tag{55}$$

and hence, *a fortiori* (cf. §4.3),

$$G_2(s) = | T(\mathcal{E}\cdot s^0) | \leqq \tfrac{1}{2}\pi L^2 | s |. \tag{56}$$

Thus obviously $T \in K_1(\mathcal{D})$ (cf. §1.32). By (55) however, it follows that T is A.C. $\mathcal{D}$ in $\mathcal{D}$—that is, T is absolutely continuous with respect to $\mathcal{D}$ itself as a base set (cf. §1.23). Hence (cf. §3.11), $N(z, T, \mathcal{D})$ is summable, and consequently the set of points z where $N(z, T, \mathcal{D}) = \infty$ is of measure zero. Consequently, for almost every point z it is true that $T^{-1}(z)$ is a finite set, and therefore every essential maximal model continuum of z under T (cf. §1.16) consists of a single point. So $N(z, T, \mathcal{E}\cdot\mathcal{D}) = \kappa(z, T, \mathcal{D})$ a.e. (cf. §1.18), and hence $T \in K_2(\mathcal{D})$ (cf. §1.34). But, as a consequence of the restricted Lipschitz condition, the partial derivatives x_u, x_v, y_u, y_v exist a.e. in $\mathcal{D}$. So $T \in K_3(\mathcal{D})$, as asserted in §1.39.

5.9. Finally, suppose that the continuous transformation T (cf. §5.1) is such that x_u, x_v, y_u, y_v exist and are continuous everywhere in $\mathcal{D}$, and that the ordinary Jacobian $J(u, v)$ is summable in $\mathcal{D}$. Take a sequence of Jordan regions $\mathfrak{R}_n$ which fill up $\mathcal{D}$ from the interior (cf. §1.13). Clearly T satisfies a restricted Lipschitz condition in $\mathfrak{R}_n^0$ for every n and hence $T \in K_3(\mathfrak{R}_n^0)$ by §5.8. The closure theorem for K_3 (cf. §1.41), applied with $T_n = T$ and $\mathcal{D}_n = \mathfrak{R}_n^0$, yields immediately the fact that $T \in K_3(\mathcal{D})$.

Bibliography

S. Banach

1. *Sur une classe de fonctions d'ensemble*, Fundamenta Mathematicae, vol. 6 (1924), pp. 170–188.
2. *Sur les lignes rectifiables et les surfaces dont l'aire est finite*, Fundamenta Mathematicae. vol. 7 (1925), pp. 225–236.

H. E. Bray

1. *Proof of a formula for an area*, Bulletin of the American Mathematical Society, vol. 29 (1923), pp. 264–270.

R. Caccioppoli

1. *Sulle funzioni additive d'insieme*, Rendiconti della R. Accademia di Scienze Fisiche e Matematiche di Napoli, (3), vol. 34 (1927), pp. 150–153.
2. *Trasformazioni piane, superficie quadrabili, integrali di superficie*, Rendiconti del Circolo Matematico di Palermo, vol. 54 (1930), pp. 217–262.

B. von Kerékjártó

1. *Vorlesungen über Topologie*, vol. 1, Berlin, 1923.

C. Kuratowski

1. *Topologie* I, Warsaw and Lwów, 1933.

E. J. McShane

1. *Integrals over surfaces in parametric form*, Annals of Mathematics, (2), vol. 34 (1933), pp. 815–838.

C. B. Morrey

1. *A class of representations of manifolds, part* I, American Journal of Mathematics, vol. 55 (1933), pp. 683–707.

H. Rademacher

1. *Über partielle und totale Differenzierbarkeit von Funktionen mehrerer Variabeln*, parts I and II, Mathematische Annalen, vol. 79 (1919), pp. 340–359, and vol. 81 (1920), pp. 52–63.

T. Radó

1. *Über das Flächenmass rektifizierbarer Flächen*, Mathematische Annalen, vol. 100 (1928), pp. 445–479.
2. *On continuous transformations in the plane*, Fundamenta Mathematicae, vol. 27 (1936), pp. 201–211.
3. *A remark on the area of surfaces*, American Journal of Mathematics, vol. 58 (1936), pp. 598–606.
4. *On the derivative of the Lebesgue area of continuous surfaces*, Fundamenta Mathematicae, vol. 30 (1938), pp. 34–39.
5. *On absolutely continuous transformations in the plane*, Duke Mathematical Journal, vol. 4 (1938), pp. 189–221.

S. Saks

1. *Theory of the Integral*, New York, Stechert, 1937.

J. Schauder

1. *Über stetige Abbildungen*, Fundamenta Mathematicae, vol. 12 (1928), pp. 47–74.

K. Szilárd

1. *Untersuchungen über die Grundlagen der Funktionentheorie*, Mathematische Zeitschrift, vol. 26 (1927), pp. 653–671.

C. de la Vallée Poussin

1. *Sur l'intégrale de Lebesgue*, these Transactions, vol. 16 (1915), pp. 435–501.

W. H. Young

1. *On a formula for an area*, Proceedings of the London Mathematical Society, vol. 18 (1920), pp. 339–374.
2. *On a new set of conditions for a formula for an area*, Proceedings of the London Mathematical Society, vol. 21 (1923), pp. 75–94.
3. *Integration over the area of a curve and the transformation of the variables in a multiple integral*, Proceedings of the London Mathematical Society, vol. 21 (1923), pp. 161–190.

The Ohio State University,
Columbus, Ohio.

ON THE CLASSIFICATION OF THE MAPPINGS OF A 2-COMPLEX

BY

HERBERT ROBBINS

1. **Introduction.** The continuous mappings of one complex into another may be classified in two essentially different ways. Two mappings are called *homotopic* if one may be deformed into the other, and the relation of homotopy partitions the set of all mappings of a complex K into a complex T into disjoint subsets of homotopic mappings. This classification is possible for the most general spaces, irrespective of their combinatorial structure. The other way of classifying mappings is according to their homology behavior. Any mapping f of K into T may be deformed into a simplicial mapping, assigning to the vertices of each simplex of K the vertices of a simplex of T of the same or lower dimension, and affine in each simplex(1). Such a mapping induces a homomorphism of the homology groups of K into those of T, and of the cohomology groups of T into those of K, for each dimension r and coefficient group G. Two mappings which induce the same homomorphisms for each r and G may be called *homologous*, and the set of mappings of K into T may be classified by this relation. It is easy to show that if two mappings are homotopic they are homologous. The converse is not true in general, but holds if K is an n-complex and T an n-sphere. This theorem is due to H. Hopf, and has been generalized by W. Hurewicz to the case where T is one of a more general class of spaces whose rth homotopy groups vanish for $r<n$(2).

Mappings may also be classified according to the homomorphisms they induce of the fundamental group of K into that of T. Again, homotopic mappings induce the same homomorphisms. Under certain conditions, when T is an "aspherical" space, whose rth homotopy groups vanish for $r>1$, these classifications coincide, as shown by Brouwer and Hurewicz(3).

Although in general homotopy provides a more finely graduated classification of mappings than combinatorial and group-theoretical methods, it is one

Presented to the Society, September 6, 1938; received by the editors May 25, 1940. This paper contains the results of a doctoral dissertation accepted by Harvard University in May, 1938.

(1) See Alexandroff-Hopf, *Topologie*, pp. 314–319.

(2) See Hopf, Commentarii Mathematici Helvetici, vol. 5 (1932), pp. 39–54, also AH, chap. 13, and Hurewicz, Proceedings, Akademie van Wetenschappen, Amsterdam, vols. 38–39 (1935–1936); in particular, vol. 39, pp. 117–126.

(3) See Brouwer, Mathematische Annalen, vol. 82 (1921), Hopf, Journal für die reine und angewandte Mathematik, vol. 165 (1931), pp. 225–236, and Hurewicz, loc. cit., vol. 39, pp. 215–222.

of the objects of combinatorial topology to give combinatorial conditions which will be sufficiently strong to ensure homotopy. In all cases where such conditions have been given heretofore, with the exception of some recent work by H. Whitney([4]), the space T is assumed to be simply-connected, or the higher dimensional homotopy groups are assumed to vanish. It is the object of this paper to give combinatorial methods which shall be adequate to classify the mappings of a 2-complex $K=K^2$ into a perfectly general space T. The results of the present paper give necessary and sufficient conditions for the homotopy of two mappings of a 2-complex into a space T; these involve the use of chains in K with coefficients from the fundamental group and 2-dim. homotopy group of T, and are given in Theorem 4.

The author wishes to express his deep indebtedness to the work of Whitney mentioned above for the stimulation of his interest in this subject. Without the help of Professor Whitney's many suggestions and constant counsel and encouragement this paper could not have been written.

2. **Coboundary and product of chains.** By the coboundary $\delta\sigma^n$ of an n-simplex $\sigma^n = x_0x_1 \cdots x_n$ we mean the set of all $(n+1)$-simplexes which have it as an n-face, with orientations determined by $\delta x_0x_1 \cdots x_n = \sum xx_0x_1 \cdots x_n$ where the summation is over all vertices x of K such that $xx_0x_1 \cdots x_n$ is an $(n+1)$-simplex of K([5]). A chain is a cocycle if its coboundary vanishes, and two cocycles are cohomologous if their difference is a coboundary.

Let the vertices of the complex K be given in a definite order, and suppose that $x_ix_j \cdots x_l$ is positively oriented when $i<j< \cdots <l$. Let $A^r=\sum\alpha_{i_0\cdots i_r}x_{i_0}\cdots x_{i_r}$ and $B^s=\sum\beta_{i_0\cdots i_s}x_{i_0}\cdots x_{i_s}$ be r- and s-chains respectively. Then by their *product* $A^r \cup B^s$ we mean the $(r+s)$-chain([6])

$$A^r \cup B^s = \sum \alpha_{i_0\cdots i_r}\beta_{i_r\cdots i_{r+s}}x_{i_0}\cdots x_{i_{r+s}}.$$

The coefficients α and β may be from a group or from any system for which an operation $\alpha\beta$ is defined.

3. **The space T and its covering space.** Let P be a fixed point of the space T([7]). The paths Pp from P to the points p of T define the points of the *covering space* $\overline{T}$. Two paths are equivalent if they have the same endpoint and if one may be deformed into the other leaving the endpoints fixed; equivalent paths define the same point of $\overline{T}$, which is topologized in the usual way, by letting a neighborhood of a path consist of all paths which are continuations

([4]) Unpublished; classifying mappings into projective spaces. Many of the methods used in this paper have been used in the present work. See also S. Eilenberg, Annals of Mathematics, vol. 41 (1940), pp. 231–251, for references to work of Pontrjagin and Freudenthal.

([5]) See Whitney, three papers in Duke Mathematical Journal, vol. 3 (1937); especially pp. 51–55.

([6]) See Whitney, Proceedings of the National Academy of Sciences, vol. 23 (1937), pp. 285–291; especially p. 286.

([7]) It is usual to assume that T is connected and locally 0- and 1-connected.

of it within a given neighborhood of its endpoint. We shall denote by Φ the function which assigns to each point of $\bar{T}$ the endpoint of a path which defines it; then Φ maps $\bar{T}$ onto T. A path (or rather, a class of equivalent paths) with endpoint p will be said to lie over the point p, and may be denoted by $\bar{p}_i$, the subscript denoting the particular path from P to p which defines it. We denote the "point-path" which lies over P by $\bar{P}$.

Let S be a simply-connected space (that is, one whose fundamental group vanishes) with a fixed point s, and let f map S into T with $f(s)=P$. Then there is a mapping $\bar{f}$ of S into $\bar{T}$ with $\bar{f}(s)=\bar{P}$, defined as follows: choose a path sx in S from s to x, and let $\bar{f}(x)$ be the path in T (i.e., point of $\bar{T}$) defined by $f(sx)$. Since S is simply-connected, this is independent of the particular path chosen. We call $\bar{f}$ the mapping *induced* by f. Clearly, $\Phi\bar{f}=f$. Conversely, let F be a mapping of S into $\bar{T}$ with $F(s)=\bar{P}$; then $f=\Phi F$ is a mapping of S into T with $f(s)=P$ such that $\bar{f}=F$. Let f_1 and f_2 be two mappings of S into T, with $f_1(s)=f_2(s)=P$; then f_1 and f_2 may be deformed into each other, keeping s at P, if and only if $\bar{f}_1$ and $\bar{f}_2$ may be deformed into each other, keeping s at $\bar{P}$. Thus the homotopy-classes of the mappings of S into T are put into one-to-one correspondence with those of S into $\bar{T}$, with s at P or $\bar{P}$ respectively.

Let S_0^2 be the unit 2-sphere in Euclidean 3-space, with a south pole P_0. The homotopy-classes of S_0^2 into T with P_0 going into P are the elements of the 2-dimensional *homotopy group* $\pi_2(T)$(8). By pushing away from P_0 it is clear that each element of $\pi_2(T)$ may be represented by a mapping in which each point of S_0^2 goes into P, with the exception of a small 2-cell whose boundary goes into P and whose interior is mapped into T as though it were a 2-sphere corresponding to S_0^2. Clearly, this patch may be pushed over S_0^2 into any position. If the element h of $\pi_2(T)$ is defined by the mapping f, we write $h=h_f$. The sum $h_{f_1}+h_{f_2}$ of two elements of $\pi_2(T)$ is defined by a mapping of S_0^2 into T which carries each point of S_0^2 into P, except for two disjoint patches, one of which is mapped by f_1 and the other by f_2. Since the position of the patches is of no consequence, $\pi_2(T)$ is seen to be an abelian group. By the preceding paragraph, $\pi_2(T)\approx\pi_2(\bar{T})$, where the isomorphism is defined by letting each mapping f correspond to its induced mapping $\bar{f}$.

Let G be the fundamental group of T, with unit element 1. We shall define an operation gh, where g is any element of G and h is in $\pi_2(T)$, such that gh is in $\pi_2(T)$ and the laws

$$g_1(g_2h) = (g_1g_2)h, \tag{3.1}$$

$$1h = h, \tag{3.2}$$

$$g(h_1 + h_2) = gh_1 + gh_2 \tag{3.3}$$

(8) For homotopy groups, see Hurewicz, loc. cit. S_0^2 may be replaced by any homeomorphic set, with a fixed point and a chosen orientation.

hold. The definition of gh_f is as follows: let a circle rise from P_0 to the equator as t goes from 0 to 1; map this 1-cell of circles into the points of a path Γ in T defining g, each circle C_t going into the corresponding point Γ_t of T, and map the upper hemisphere of S_0^2 into T by f. The resulting mapping defines gh_f, and the three properties stated are easily verified.

If a fixed mapping f defining h_f of $\pi_2(T)$ carries another point P^* into P also, we may regard P^* as the fixed point of S_0^2 and get another element h_{f^*}. More precisely, we may consider the mapping f^* of S_0^2 into T obtained by first rotating S_0^2 to bring P^* to P_0 and then mapping by f. Suppose f maps a path P_0P^* into g of G. Then

$$h_{f^*} = g^{-1}h_f. \tag{3.4}$$

This is clear if we consider the induced mappings $\bar{f}$ and $\bar{f}^*$ into $\overline{T}$. For under $\bar{f}$, a point x of S_0^2 is mapped into a path $f(P_0x)$, while under $\bar{f}^*$ it is mapped into a path which is equivalent to the path $g^{-1}f(P_0x)$. This fact may be used to give a simple geometrical proof of (3.3).

Suppose f maps S_0^2 into T, defining the element h of $\pi_2(T)$, in the following manner: a small 2-cell σ goes into P, and a path P_0P^* from P_0 to the boundary of σ goes into the element g of G. Then if we replace the mapping of σ by one which (regarding σ as a 2-sphere) defines the element h' of $\pi_2(T)$, the resulting mapping of the whole of S_0^2 defines an element H of $\pi_2(T)$, and

$$H = h + gh'. \tag{3.5}$$

For regarding P^* as the fixed point of S_0^2 shows that $g^{-1}H = g^{-1}h + h'$ by definition of the sum of two elements of $\pi_2(T)$, and going back to P_0 (i.e., applying g) gives the desired equality.

We shall have occasion to regard the boundary ∂E^3 of a 3-cell E^3 as a 2-sphere, and a mapping of ∂E^3 as defining an element of $\pi_2(T)$. This is completely specified as soon as we have chosen a particular point of ∂E^3 as the fixed point; the particular homeomorphism of ∂E^3 with S_0^2 which defines the element of $\pi_2(T)$ may be set up by placing the 3-cell inside S_0^2 and projecting from the center of S_0^2 so that the point chosen as the fixed point of ∂E^3 goes into P_0. The precise manner in which this is done is immaterial, since we are concerned only with homotopy invariants.

Suppose $E_1^3 = x_0x_1x_2x_3$ and $E_2^3 = y_0y_1y_2y_3$ are two 3-cells whose boundaries are regarded as 2-spheres with fixed points x_0 and y_0 respectively, and suppose f_1 and f_2 map E_1^3 and E_2^3 into T with $f_1(x_0) = f_2(y_0) = P$. Then two elements h_{f_1} and h_{f_2} of $\pi_2(T)$ are defined. Now suppose that the 2-cells $x_0x_1x_2$ and $y_0y_1y_2$ are congruent, with x_i corresponding to y_i, and that f_1 on $x_0x_1x_2$ coincides with f_2 on $y_0y_1y_2$ under this congruence. Then we may form a 2-sphere with fixed point x_0 by placing E_1^3 and E_2^3 together, bringing x_i into coincidence with y_i ($i = 0, 1, 2$) and dropping out the 2-cells $x_0x_1x_2$ and $y_0y_1y_2$. Denoting the element of $\pi_2(T)$ thus defined by h, we have(9)

$$h = h_{f_1} + h_{f_2}. \tag{3.6}$$

For $h_{f_1}+h_{f_2}$ is defined by a mapping homotopic to that defined by h, except that there are two 2-cells adjacent along an edge, and such that the mapping of one is the reflection of the mapping of the other in this edge. This patch may be eliminated by a simple deformation, which proves the above relation. (3.6) also holds if the 3-simplexes E_i^3 are replaced by 3-cells, with mappings coinciding along congruent 2-faces.

Let G be a group (not necessarily abelian) with elements g and operation $g_1g_2=g_3$. By the ***group ring*** $\overline{G}$ we mean the set of linear forms $\sum a_ig_i$, where the a_i are integers, and only a finite number of terms appear. We assume the laws

$$(a + b)g = ag + bg, \tag{3.7}$$

$$a(bg) = (ab)g. \tag{3.8}$$

Then $\overline{G}$ is a ring, where $+$ is defined as formal addition, and $\cdot$ by the law[10]

$$\sum a_ig_i \cdot \sum b_ig_i^* = \sum (a_ib_j)g_ig_j^*. \tag{3.9}$$

We now assume that G operates on the abelian group H, i.e., that a multiplication gh is defined, where gh is in H, and the laws (3.1) to (3.3) are satisfied. If we now define

$$\left(\sum a_ig_i\right)h = \sum a_i(g_ih), \tag{3.10}$$

then $\overline{G}$ will be a ring of operators on H, the laws (3.1) to (3.3) remaining valid. In future applications, G will be the fundamental group of T and H will be $\pi_2(T)$.

Let the fundamental group G of T be given by a set of generators $g_1, g_2, \cdots$ and a set of relations $R_1, R_2, \cdots$. Any product of elements of G may be called a word, and any word may be written as a product $g_1^{\alpha_1}g_2^{\alpha_2} \cdots g_p^{\alpha_p}$ of the generators. For each element g of G we choose a fixed representation as a product of the g_i, and call this the ***normal form*** for g. Any product of the g_i which equals 1 may be shown to do so by using a succession of the relations R_i, together with the trivial relations $g_ig_i^{-1}=g_i^{-1}g_i=1$; for each product we choose a definite manner of doing so. For each generator g_i we choose a definite path σ_i in T determining it; if $g=g_1^{\alpha_1} \cdots g_p^{\alpha_p}$ in normal form, we choose as the definite path determining it the path $\sigma_1^{\alpha_1} \cdots \sigma_p^{\alpha_p}$. Then each relation says that a certain path may be shrunk to a point. If the path is $\theta=\sigma_1^{\alpha_1} \cdots \sigma_p^{\alpha_p}$, then if the 1-sphere S' is mapped over θ, this mapping may be extended throughout the interior R^2-S' of S'. We choose a definite manner of doing this; then to each R_j corresponds a mapping ϕ_j of R^2 into T, defining the (singular) 2-chain

(9) We suppose the 2-spheres are oriented like $\partial(x_0x_1x_2x_3)$, $-\partial(y_0y_1y_2y_3)$, and $\partial(x_0x_1x_2x_3 - y_0y_1y_2y_3)$.

(10) See K. Reidemeister, Mathematische Annalen, vol. 112 (1936) for a similar use of $\overline{G}$.

$C_i=\phi_i(R^2)$ of T, with boundary $\sum\alpha_i\sigma_i$. Likewise there exist the induced mappings $\bar{\phi}_i$ of R^2 into $\bar{T}$, and the chains $\bar{C}_i=\bar{\phi}_i(R^2)$.

4. **The complex K; standard mappings.** Let K be a connected, simplicial complex with vertices x_i which are ordered in a definite manner according to their subscripts: $x_1<x_2<x_3< \cdots$. We assume that the simplexes $x_ix_j\cdots x_l$ are positively oriented if $i<j<\cdots<l$.

Let f map K into T. We call f *normal over* K^1—the 1-dim. part of K—if it maps each vertex x_i of K into the fixed point P of T, and each 1-cell x_ix_j into an element $\sigma_1^{\alpha_1}\cdots\sigma_r^{\alpha_r}$ of G in normal form, with the convention that the normal form for the path defining the unit element of G is the "point-path" P.

LEMMA 1. *Any mapping f of K into T may be deformed into a mapping normal over K^1.*

The required deformation is obtained by first bringing all the vertices to P, then extending the deformation through the rest of K, then deforming the mapping so that all the 1-cells are in normal form and extending the deformation through K, using a simple lemma on such extensions(11).

Let f be a mapping of K into T, normal over K^1. Then we may define a mapping f^0 of K''(12) into T, coinciding with f on K^1, and defined in the 2-cells of K as follows: running around the boundary of a 2-cell σ^2 of K defines three elements of G, each in normal form; there is a corresponding definite deformation of their product to P (since this is equivalent to the existence of an extension of the mapping defined on the boundary of σ^2 throughout its interior) using the R_i, and corresponding mappings into T. Thus we define f^0 throughout the interior of σ^2, giving $f^0(\sigma^2)=\sum\alpha_iC_i$. A mapping f which coincides with the mapping f^0 thus determined throughout a 2-cell σ^2 we shall call *standard over* σ^2; a mapping which is standard over each 2-cell of K we shall call *standard* (note that if $\sigma=x_ix_jx_k$, $i<j<k$, the three elements of g are, in order, $g(x_ix_j)$, $g(x_jx_k)$, and $g^{-1}(x_ix_k)$.)

5. **An example.** In this section we shall illustrate the concepts defined in the preceding pages by choosing a particular space $T=$"torus with patch," defined as follows: choose a definite simplicial subdivision of the torus, and adjoin to it another 2-simplex A whose boundary only is identified with that of a congruent 2-simplex of the torus. Intuitively, this space corresponds to an inflated inner tube with a small patch cemented to the tube around its edge. The fundamental group G of this space has two generators, g_1 and g_2, with the single relation

$$\text{(R)} \qquad g_1g_2 = g_2g_1.$$

The elements of G may be represented by ordered pairs of integers (α_1, α_2), with the law of addition given by $(\alpha_1, \alpha_2) = (\beta_1, \beta_2) = (\alpha_1+\beta_1, \alpha_2+\beta_2)$.

(11) See Whitney, loc. cit., in note 5, pp. 52–53.

(12) K'' is the 2-dim. part of K.

We choose paths σ_i defining the g_i as two circles on the torus; then corresponding to the relation (R) we have the fact that if the boundary of a 2-cell σ^2 be mapped into the path $\sigma_1+\sigma_2-\sigma_1-\sigma_2$, this mapping may be extended throughout the interior of σ^2. It is natural to choose as the corresponding definite mapping of the interior of the 2-cell into T the mapping into all of T except the patch; i.e., into $T-A$, as though T were simply a torus.

Each element of G is of the form (α_1, α_2), where the α_i are positive, negative, or zero; we choose as the normal form for such an element the path which

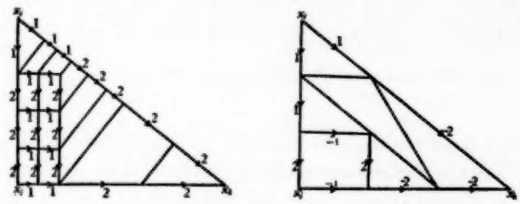


FIG. 1

first runs around $\sigma_1\alpha_1$ times, then around $\sigma_2\alpha_2$ times (in the positive or negative senses, according to the signs of the α_i).

If f is a normal mapping of the boundary of a 2-cell $x_ix_jx_k$ into T for which

$$g(x_ix_j) = (\alpha_1, \alpha_2), \qquad g(x_jx_k) = (\beta_1, \beta_2), \qquad g(x_ix_k) = (\gamma_1, \gamma_2)$$

are the elements of G into which the 1-cells of $x_ix_jx_k$ are mapped by f, then the necessary and sufficient condition that f may be extended throughout the interior of $x_ix_jx_k$ is that

$$\alpha_1 + \beta_1 = \gamma_1, \qquad \alpha_2 + \beta_2 = \gamma_2.$$

(This will be the case if f is derived from any mapping of K into T normal over K^1, and $x_ix_jx_k$ is a 2-cell of K.) As our definite manner of deforming the boundary of $x_ix_jx_k$ to P, we choose the following sequence of equalities in G, using the trivial relations and (R):

$$\begin{aligned}\alpha_1\sigma_1 + \alpha_2\sigma_2 + \beta_1\sigma_1 + \beta_2\sigma_2 &= \alpha_1\sigma_1 + (\alpha_2 - 1)\sigma_2 + \sigma_2 + \sigma_1 + (\beta_1 - 1)\sigma_1 + \beta_2\sigma_2 \\ &\to \alpha_1\sigma_1 + (\alpha_2 - 1)\sigma_2 + \sigma_1 + \sigma_2 + (\beta_1 - 1)\sigma_1 + \beta_2\sigma_2 \\ &\to \cdots \to (\alpha_1 + \beta_1)\sigma_1 + (\alpha_2 + \beta_2)\sigma_2 = \gamma_1\sigma_1 + \gamma_2\sigma_2.\end{aligned}$$

In Figure 1 we indicate the corresponding mappings of the interior of $x_ix_jx_k$ into T for two cases:

$$(\alpha_1, \alpha_2) = (1, 3), \quad (\beta_1, \beta_2) = (2, 2); \quad (\alpha_1, \alpha_2) = (2, 1), \quad (\beta_1, \beta_2) = (-1, -2).$$

The general situation is then clear. Note that in the resulting subdivision of $x_ix_jx_k$, each small 2-cell is mapped either into $\pm\sigma_i$ or into $T-A$, but in no

case does a point of $x_ix_jx_k$ lie in A. This gives the definition of standard mapping for this example.

The covering space $\overline{T}$ of T is the plane, ruled off into congruent rectangles as in the case of the torus alone, except that here each rectangle has a small patch like that of T. We may number each rectangle with a double subscript, and denote the parts of each rectangle by the symbol denoting the corresponding part of T, with a bar above and a double subscript:

$$\overline{T}_{ij} \to T, \qquad \overline{A}_{ij} \to A, \qquad \overline{P}_{ij} \to P.^{(12)}$$

The nature of $\pi_2(T)$ follows from the

LEMMA 2. *Let W be a simply-connected space with fixed point P, and let S^2 be a 2-sphere. By $W+S^2$ we mean the space obtained by identifying a single point of S^2 with P. Then $\pi_2(W+S^2)$ is isomorphic with $\pi_2(W)+I_0$, where I_0 is the additive group of integers.*

Proof. Let f be a mapping of the 2-sphere S_0^2 subdivided simplicially into $W+S^2$ with $f(P_0)=P$. The subdivision may be chosen so that each 2-cell σ_i^2 of S_0^2 is mapped into either W or S^2 but not both. We may deform f so that all vertices of S_0^2 lie at P, and since $W+S^2$ is simply-connected, we may further deform f so that all the 1-cells of S_0^2 lie at P. A further deformation will now be made. Consider any 2-cell σ^2 of S_0^2; its boundary is mapped into P. We deform f in σ^2 by shrinking σ^2 into a smaller 2-cell lying within the original σ^2, mapping this small 2-cell just as f mapped σ^2, and mapping the region between the original boundary of σ^2 and the new 2-cell into P. Clearly, this may be done by a deformation of f in σ^2. Proceeding in this way we deform f in all the 2-cells of S_0^2, so that all of S_0^2 is mapped into P, except for a number of small islands, each of which is mapped either into W or into S^2. We may now push all the islands which go into W around S_0^2 so that they lie in one hemisphere, and all those which go into S^2 so that they lie in the other, and so that the equatorial circle which separates the two classes goes through P_0 and is mapped into P. The resulting mapping defines an element of $\pi_2(W)$ and one of $\pi_2(S^2)$, which is isomorphic with I_0. Thus, to each mapping f defining an element of $\pi_2(W+S^2)$ corresponds a pair of mappings f_1, f_2 of $\pi_2(W)$ and $\pi_2(S^2)$ respectively, and this correspondence is the desired isomorphism. This completes the proof.

Now consider a mapping f of S_0^2 into T = torus with patch, and the corresponding induced mapping $\bar{f}$ of S_0^2 into $\overline{T}$ = the covering space of T. Since S_0^2 is compact, we can drop all but a finite block of fundamental regions $\overline{T}_{ij}$ from $\overline{T}$ without affecting f. The remaining space may be constructed by the addition of fundamental regions to each other a finite number of times, and the lemma gives the result expressed in

(12) We remark that $\pi_2(T)$ and $\pi_2(\overline{T})$ are not altered if we pull the boundary of the patches to single points, so that each becomes a 2-sphere touching the rest of the space at a single point.

THEOREM 1. *The 2-dim. homotopy group $\pi_2(T)$, where T = torus with patch, is isomorphic with the direct sum $\sum^\infty I_0$, where the elements of $\sum^\infty I_0$ are infinite sequences of integers, only a finite number of which differ from* 0.

To bring out clearly the effect of G operating on $\pi_2(T)$ we may represent the elements of G by ordered pairs of integers $g=(\alpha_1, \alpha_2)$, and the elements of $\pi_2(T)$ by matrices $h=\|a_{ij}\|$; then gh is the matrix obtained from h by a shift of α_1 places horizontally and α_2 places vertically.

An intuitive interpretation of the element a_{ij} in the matrix of the element h of $\pi_2(T)$ corresponding to the mapping $\bar{f}$ of S_0^2 into $\bar{T}$ may be given as follows: only a finite number of fundamental regions $\bar{T}_{ij}$ will contain images of points of S_0^2. Cut $\bar{T}$ along a rectangle containing all of these, and draw rays from the center of the patch $\bar{A}_{ij}$. By pushing along these rays we may alter $\bar{f}$ so that all points of S_0^2 go into points of the little 2-sphere formed by the patch $\bar{A}_{ij}$ and its underlying simplex, while all points which were originally mapped into points of this 2-sphere remain fixed. (For this purpose we may identify all patches other than $\bar{A}_{ij}$ with their underlying simplexes.) Then we have a mapping of S_0^2 into a 2-sphere, and its degree in the ordinary sense will be precisely a_{ij}. This may be interpreted roughly as the number of times S_0^2 covers the small 2-sphere, or, equivalently, as the number of times the patch $\bar{A}_{ij}$ is covered by the mapping $\bar{f}$ of S_0^2.

The reader may find it profitable to consider the details, similar to those we have given, for the case where T is the topological product of three circles, or where the fundamental group of T is isomorphic with the additive group of integers modulo n.

6. **Normal mappings; the degree of a mapping.** We return to the case of an arbitrary space T.

The first vertex of a cell of a complex whose vertices are ordered as in §4 will be called its *leading vertex*. It is not true that any mapping of a complex K into T may be deformed into a standard mapping. But we may alter standard mappings on each 2-cell near its leading vertex so that the resulting class of mappings is perfectly general in the sense of homotopy. We do this as follows: let f^0 be a standard mapping of $x_ix_jx_k$ into T. Deform f^0, keeping the boundary fixed, so that a small 2-cell inside $x_ix_jx_k$ and touching the boundary only at x_i goes into P. Now replace f^0 on this 2-cell by a mapping which defines an arbitrary element of $\pi_2(T)$. The resulting mapping of $x_ix_jx_k$ into T will be called *normal over* $x_ix_jx_k$. If a mapping is normal over all the 2-cells of K it will be called *normal*. The following lemma holds:

LEMMA 3. *Any mapping f of K into T may be deformed into a normal mapping.*

Proof. First we may deform f into a mapping normal over K^1, by Lemma 1. Now consider any 2-cell σ^2 of K. Make another copy of σ^2 and define on it

the standard mapping f^0 which coincides with f along the boundary of σ^2, altered so that a small 2-cell near x_i goes into P, as above. Join the two 2-cells along their boundaries, and consider the resulting mapping ϕ of the 2-sphere thus formed into T. The boundary of the small 2-cell divides the 2-sphere into two parts, and itself goes into P, so that the element of $\pi_2(T)$ thus defined, *taking* x_i *as the fixed point of the* 2-*sphere*, is the sum of two elements, one of which may be chosen at pleasure. By choosing it properly, we can make the resulting mapping homotopic to 0. This is equivalent to saying that ϕ may be extended throughout the interior of the 2-sphere, or that the mapping f may be deformed into a mapping normal on σ^2, leaving the boundary of σ^2 fixed. Thus we may deform f over each 2-cell of K until it is normal.

The element of $\pi_2(T)$ by which the standard mapping f^0 must be altered to give a mapping deformable into f on $x_ix_jx_k$ will be called the *degree* of f on $x_ix_jx_k$, and denoted by $d_f(x_ix_jx_k)$. This is simply the negative of the element of $\pi_2(T)$ defined by f^0 unchanged and f when the boundaries are identified as above. Clearly, if f is already standard on $x_ix_jx_k$, then $d_f(x_ix_jx_k)=0$, so that the degree of a mapping on a 2-cell is a measure of its deviation from the corresponding standard mapping.

Let $E^3=x_ix_jx_kx_l$ be a three-cell of K, and let f map the boundary $\partial x_ix_jx_kx_l$ of E^3 into T, with the leading vertex x_i going into P. Then f defines an element of $\pi_2(T)$ if we regard ∂E^3 as a 2-sphere with fixed point x_i. We shall denote this element of $\pi_2(T)$ by $D_f(\partial E^3)$ and call it the *degree of f on ∂E^3*. Suppose f is normal over ∂E^3 and f^0 is the corresponding standard mapping. We wish to find a relation between $D_{f^0}(\partial E^3)$ and $D_f(\partial E^3)$. The mapping f is obtained from f^0 by replacing in each 2-cell of ∂E^3 a small piece going into P by one which defines an element of $\pi_2(T)$. The following theorem is an immediate consequence of (3.5):

THEOREM 2. *If $E^3=x_ix_jx_kx_l$ is a 3-cell of K, and f is a normal mapping of ∂E^3 into T, then*

$$\begin{aligned} D_f(\partial E^3) &= D_{f^0}(\partial E^3) - d_f(x_ix_jx_k) + d_f(x_ix_jx_l) \\ &\quad - d_f(x_ix_kx_l) + g(x_ix_j)d_f(x_jx_kx_l), \end{aligned} \tag{6.1}$$

where $g(x_ix_j)$ is the element of G into which f maps x_ix_j.

7. **An extension theorem.** Let f be a normal mapping of the subcomplex K^* of a simplicial 3-complex K into T; that is, all vertices x_i go into P, all 1-cells x_ix_j into elements of G in normal form, and all 2-cells of K^* are mapped normally. We wish to find conditions that f may be extended throughout all of K. First, when may f be extended throughout all the 1- and 2-cells of K? Let $\sigma^2=x_ix_jx_k$ be a 2-cell of K; then if x_ix_j, x_jx_k, and x_ix_k are mapped into the elements $g(x_ix_j)$, etc., of G, the necessary and sufficient condition that this mapping may be extended throughout the interior of σ^2 is that $g(x_ix_j)\cdot g(x_jx_k)$

$\cdot g^{-1}(x_i x_k) = 1$, or in other words, that running around the boundary of σ^2 defines the unit element of G. It is easily seen that neither the sense nor the initial point of the circuit matters for this purpose. Thus, if we define the *coboundary* δA^1 of the 1-G-chain

$$A^1 = \sum g(x_i x_j) x_i x_j \tag{7.1}$$

to be the 2-G-chain

$$A^2 = \sum g(x_i x_j x_k) x_i x_j x_k, \tag{7.2}$$

where

$$g(x_i x_j x_k) = g(x_i x_j) \cdot g(x_j x_k) \cdot g^{-1}(x_i x_k), \tag{7.3}$$

then the necessary and sufficient condition that f defined on K^* may be extended throughout all the 2-cells of K is that a 1-G-chain A_1^1 of K of the form (7.1) exist such that δA_1^1 is a 2-G-chain each of whose coefficients is 1, where $g(x_i x_j)$ is defined by f on those 1-cells $x_i x_j$ which belong to K^*. A 1-G-chain of K whose coboundary vanishes in this sense will be called a *cocycle*; then our condition is that the 1-G-chain

$$A_0^1 = \sum g(x_i x_j) x_i x_j \tag{7.4}$$

summed over all $x_i x_j$ in K^* be *part of a cocycle* A_1^1; i.e., that elements $g(x_i x_j)$ may be assigned to the 1-cells $x_i x_j$ in $K - K^*$ so that the chain (7.4) then becomes a cocycle. We shall return to this matter of chains with coefficients from a non-abelian group in §9.

Now we assume that f has been defined throughout all the 2-cells of K; when may it be extended throughout all the 3-cells? Let $E^3 = x_i x_j x_k x_l$ be such a 3-cell; regard its boundary as a 2-sphere with fixed point x_i. Then f may be extended throughout the interior of E^3 if and only if

$$D_f(\partial E^3) = 0, \tag{7.5}$$

and f may be extended throughout all the 3-cells of K if and only if $C^3 = \sum D_f(\partial E^3) E^3 = 0$, where the summation is over all 3-cells $x_i x_j x_k x_l$ of K. By equation (6.1) we have

$$\begin{aligned} C^3 = \sum D_f(\partial x_i x_j x_k x_l) x_i x_j x_k x_l + \sum [&- d_f(x_i x_j x_k) + d_f(x_i x_j x_l) \\ &- d_f(x_i x_k x_l) + g(x_i x_j) d_f(x_j x_k x_l)] x_i x_j x_k x_l. \end{aligned} \tag{7.6}$$

We shall now consider the meaning of the first term on the right side of this equation.

As usual, let G be the fundamental group of T and $\pi_2(T)$ the 2-dim. homotopy group. We shall define a new sort of operation $g_1 \circ g_2$ on the elements of G giving standard mappings of 2-cells into T, and an operation $g_1 \circ g_2 \circ g_3$ giving elements of $\pi_2(T)$:

(a) $g_1 \circ g_2$ is the standard mapping of a 2-cell $x_ix_jx_k$ into T defined as follows: map x_ix_j into g_1 and x_jx_k into g_2; then the uniquely determined standard mapping of $x_ix_jx_k$ into T will be denoted by $g_1 \circ g_2$.

(b) Let $\phi = g_1 \circ g_2$ be a standard mapping of $x_ix_jx_k$ into T; then by $\phi \circ g_3 = (g_1 \circ g_2) \circ g_3$ we mean the standard mapping of $\partial x_ix_jx_kx_l$ into T determined by mapping $x_ix_jx_k$ into T by ϕ and x_kx_l by g_3. We may omit parentheses and define $g_1 \circ g_2 \circ g_3$ to be the standard mapping of $\partial x_ix_jx_kx_l$ into T defined by mapping x_ix_j, x_jx_k, and x_kx_l into g_1, g_2, and g_3 respectively. This mapping is uniquely determined. For knowing $g(x_ix_j)$ and $g(x_jx_k)$ gives $g(x_ix_k)$, defining uniquely the standard mapping on $x_ix_jx_k$, and likewise for the other faces of $x_ix_jx_kx_l$. Since each standard mapping of $\partial x_ix_jx_kx_l$ into T determines uniquely an element of $\pi_2(T)$ we may regard the operation $\circ$ when applied to three elements of G as giving an element of $\pi_2(T)$[14].

Now let us return to (7.6). As before, let

$$A_1^1 = \sum g(x_ix_j)x_ix_j, \tag{7.7}$$

summed over all x_ix_j of K be the cocycle of which A_0^1 is a part; then

$$\sum D_f(\partial x_ix_jx_kx_l)x_ix_jx_kx_l = A_1^1 \cup A_1^1 \cup A_1^1, \tag{7.8}$$

where the multiplication of elements of G is understood in the sense just defined.

It remains to investigate the second sum of (7.6). We may write

$$\begin{aligned} &- d_f(x_ix_jx_k) + d_f(x_ix_jx_l) - d_f(x_ix_kx_l) + g(x_ix_j)d_f(x_jx_kx_l) \\ &= - d_f(x_ix_jx_k) + d_f(x_ix_jx_l) - d_f(x_ix_kx_l) + d_f(x_jx_kx_l) \\ &\quad + (g(x_ix_j) - 1)d_f(x_jx_kx_l), \end{aligned} \tag{7.9}$$

where $(g(x_ix_j) - 1)$ is an element of $\overline{G}$. Let

$$A_1^2 = \sum d_f(x_ix_jx_k)x_ix_jx_k, \tag{7.10}$$

$$A_3^1 = \sum (g(x_ix_j) - 1)x_ix_j, \tag{7.11}$$

where the summation is over all 2- and 1-cells of K, respectively. Then from (7.6), (7.8), and (7.9) we have

$$C^3 = A_1^1 \cup A_1^1 \cup A_1^1 + \delta A_1^2 + A_3^1 \cup A_1^2. \tag{7.12}$$

This proves

THEOREM 3. *Let K^* be a subcomplex of $K = K^3$ and let f be a normal mapping*

---

(14) Perhaps the main difficulty of the classification problem is that of determining this multiplication in any concrete case.

of K^* into T. Necessary and sufficient conditions that f may be extended throughout K are

(i) *that A_0^1 be part of a cocycle A_1^1,*

(ii) *that the 2-chain A_1^2 with coefficients from $\pi_2(T)$ exist such that $A_1^1 \cup A_1^1 \cup A_1^1 + \delta A_1^2 + A_3^1 \cup A_1^2 = 0$,*

where

$$A_0^1 = \sum g(x_i x_j) x_i x_j$$

summed over all $x_i x_j$ of K^*,

$$A_1^2 = \sum d_{ijk} x_i x_j x_k$$

summed over all $x_i x_j x_k$ of K and such that $d_{ijk} = d_f(x_i x_j x_k)$ whenever $x_i x_j x_k$ is in K^*, and

$$A_3^1 = \sum (g(x_i x_j) - 1) x_i x_j = A_1^1 - I^1$$

where $I^1 = \sum 1 \cdot x_i x_j$ summed over all $x_i x_j$ of K.

Note that in $A_1^1 \cup A_1^1 \cup A_1^1$ the "product" of elements of G is either a standard mapping or an element of $\pi_2(T)$, while in $A_3^1 \cup A_1^2$ the elements of G are operators on $\pi_2(T)$.

8. **The classification.** Let K be a 2-complex and let f^1 and f^2 be normal mappings of K into the space T. Let $K \times I$ be the product-complex of K with the unit interval; then f^1 and f^2 are homotopic if and only if there exists a mapping F of $K \times I$ into T such that $F(x, 0) \equiv f^1(x)$ and $F(x, 1) \equiv f^2(x)$. In this section we shall specialize the result of §7 to the case where $K^3 = K^2 \times I$, and where F is defined on $K \times 0 + K \times 1$ by f^1 and f^2 respectively.

We may subdivide $K \times I$ into cells of the form $\sigma_j^i \times I$, where the σ_j^i are the cells of K. If in particular K is of the form considered in §4, we shall orient $K \times I$ as follows: the orientation of $K \times 0$ and $K \times 1$ shall be as in K, with corresponding vertices y_i and z_i. The 1-cells $y_i z_i$ shall be positively oriented in that form. Orientation of the 2-cells $x_i x_j \times I$ shall be such as to put $y_i y_j$ on the boundary of $x_i x_j \times I$ in the positive sense, and the oriented 3-cell $x_i x_j x_k \times I$ shall have $y_i y_j y_k$ on its boundary in the positive sense. Of course, $K \times I$ as defined is not a simplicial complex; it may be simplicially subdivided, and we shall do this later.

When may F as defined above be extended throughout all the 2-cells of $K \times I$? The remaining 2-cells are of the form $x_i x_j \times I$. Clearly, the necessary and sufficient condition is that elements a_i of G exist such that

$$f_{ij}^1 a_j (f_{ij}^2)^{-1} (a_i)^{-1} = 1 \tag{8.1}$$

for all $i < j$ such that $x_i x_j$ is a 1-cell of K, and where f_{ij}^1 and f_{ij}^2 are respectively the elements of G into which f^1 and f^2 map $x_i x_j$. We may write this in the form

$$f^{1}_{ij} = a_i f^{2}_{ij} a_j^{-1}. \tag{8.2}$$

We shall call two 1-G-chains $C^1 = \sum g_{ij} x_i x_j$ and $C^2 = \sum h_{ij} x_i x_j$ *cohomologous* if elements α_i of G exist such that

$$g_{ij} = \alpha_i h_{ij} \alpha_j^{-1}, \tag{8.3}$$

and write $C^1 \backsim C^2$; this reduces to the ordinary definition if G is abelian. It is natural to write $A_1^1/A_2^1 = \delta A^0$ if A_1^1 and A_2^1 are so related, where

$$A^0 = \sum \alpha_i x_i. \tag{8.4}$$

The 0-G-chain A^0 is uniquely determined, if K is connected, when the coefficient of any vertex is assigned.

We may now assert that F as defined may be extended throughout all the 2-cells of $K \times I$ if and only if

$$A_1^1 \backsim A_2^1 \quad \text{or} \quad A_1^1/A_2^1 = \delta A^0 \tag{8.5}$$

where

$$A_1^1 = \sum f^{1}_{ij} x_i x_j, \qquad A_2^1 = \sum f^{2}_{ij} x_i x_j, \qquad A^0 = \sum a_i x_i. \tag{8.6}$$

Let us now suppose the extension throughout the 2-cells $x_i x_j \times I$ made in a normal manner; when may we extend the resulting mapping throughout the 3-cells $x_i x_j x_k \times I$? The boundary of the 3-cell $E^3 = x_i x_j x_k \times I$ may be regarded as a 2-sphere *with fixed point* y_i, and the normal mapping of it defines an element $D_F(\partial E^3)$. The necessary and sufficient condition that this mapping may be extended throughout the interior of E^3 is that $D_F(\partial E^3) = 0$, from which it follows that the n. and s. condition that this be possible for all the 3-cells of $K \times I$ is that

$$\sum D_F(\partial E^3) E^3 = \sum D_F(\partial x_i x_j x_k \times I) x_i x_j x_k \times I = 0. \tag{8.7}$$

It follows from (3.5) that

$$\begin{aligned} D_F(\partial E^3) = D_{F^0}(\partial E^3) &+ d_{f^1}(x_i x_j x_k) - d_F(x_i x_j \times I) \\ &+ d_F(x_i x_k \times I) - f^{1}_{ij} d_F(x_j x_k \times I) - a_i d_{f^2}(x_i x_j x_k). \end{aligned} \tag{8.8}$$

Now let

$$\begin{aligned} \eta_{ijk} = d_{f^1}(x_i x_j x_k) - a_i d_{f^2}(x_i x_j x_k), \qquad &\zeta_{ijk} = D_{F^0}\partial(x_i x_j x_k \times I), \\ b_{ijk} = d_F(x_i x_j \times I). \end{aligned} \tag{8.9}$$

Then

$$D_F(\partial x_i x_j x_k \times I) = \zeta_{ijk} + \eta_{ijk} - [b_{ij} - b_{ik} + b_{jk}] - (f^{1}_{ij} - 1) b_{ik}, \tag{8.10}$$

so that

$$\begin{aligned}\sum D_F(\partial E^3)E^3 = \sum \zeta_{ijk}x_ix_jx_k \times I &+ \sum \eta_{ijk}x_ix_jx_k \times I \\ (8.11) \qquad - \sum &[b_{ij} - b_{ik} + b_{jk}]x_ix_jx_k \times I \\ - \sum &(\overset{1}{f}_{ij} - 1)b_{jk}x_ix_jx_k \times I.\end{aligned}$$

Now let

$$(8.12) \qquad \begin{aligned} \overset{2}{B} &= \sum \zeta_{ijk}x_ix_jx_k, & \overset{1}{B} &= \sum b_{ij}x_ix_j, \\ \overset{2}{C} &= \sum \eta_{ijk}x_ix_jx_k, & \overline{A}_1^1 &= \sum (\overset{1}{f}_{ij} - 1)x_ix_j. \end{aligned}$$

Then (8.7) becomes

$$(8.13) \qquad \overset{2}{B} + \overset{2}{C} = \delta \overset{1}{B} + \overline{A}_1^1 \cup \overset{1}{B}.$$

We shall now find a simple expression for the term B^2.

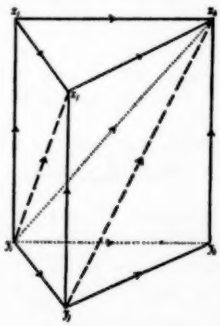


FIG. 2

We subdivide $x_ix_jx_k \times I$ simplicially, as shown in Figure 2. (This subdivision is easy to define combinatorially.) Then

$$(8.14) \qquad \begin{aligned} A_1^1 \cup A_1^1 \cup \overset{0}{A} &= \sum (\overset{1}{f}_{ij} \circ \overset{1}{f}_{jk} \circ a_k)x_ix_jx_k, \\ A_1^1 \cup \overset{0}{A} \cup A_2^1 &= \sum (\overset{1}{f}_{ij} \circ a_j \circ \overset{2}{f}_{jk})x_ix_jx_k, \\ \overset{0}{A} \cup A_2^1 \cup A_2^1 &= \sum (a_i \circ \overset{2}{f}_{ij} \circ \overset{2}{f}_{jk})x_ix_jx_k, \end{aligned}$$

and from (3.6) it follows, taking account of orientation, that

$$(8.15) \qquad \overset{2}{B} = - A_1^1 \cup A_1^1 \cup \overset{0}{A} + A_1^1 \cup \overset{0}{A} \cup A_2^1 - \overset{0}{A} \cup A_2^1 \cup A_2^1.$$

Thus (8.13) becomes

$$(8.16) \qquad \begin{aligned} \overset{2}{C} - A_1^1 \cup A_1^1 \cup \overset{0}{A} + A_1^1 \cup \overset{0}{A} \cup A_2^1 - \overset{0}{A} \cup A_2^1 \cup A_2^1 \\ = \delta \overset{1}{B} + \overline{A}_1^1 \cup \overset{1}{B}. \end{aligned}$$

We may now state

THEOREM 4. *A necessary and sufficient condition that the normal mappings* f^1 *and* f^2 *of* $K=K^2$ *into* T *be homotopic is that a 0-G-chain* A^0 *and a* 1-$\pi_2(T)$-*chain* B^1 *of* K *exist, satisfying*

(i) $A_1^1/A_2^1=\delta A^0$,

(ii) (8.16),

where

$$A_1^1 = \sum f_{ij}^1 x_i x_j, \qquad A_2^1 = \sum f_{ij}^2 x_i x_j, \qquad \overline{A}_1^1 = \sum (f_{ij}^1 - 1) x_i x_j = A_1^1 - I^1,$$
$$C^2 = \sum d_{f^1}(x_i x_j x_k) - a_i d_{f^2}(x_i x_j x_k) x_i x_j x_k,$$

and where $A^0=\sum a_i x_i$.

In certain cases, (8.16) becomes simplified, since the chain B^2 is automatically 0; this is so when T = torus with patch, for the standard mappings of ∂E^2 into T cover no patches, and are therefore of degree 0, since the 2-dim. homotopy group of the plane vanishes. For this case, (8.16) becomes simply

$$C^2 = \delta B^1 + \overline{A}_1^1 \cup B^1. \tag{8.17}$$

This is also the case for any space with the fundamental group of the torus; of course the interpretation of the elements of $\pi_2(T)$ as matrices of integers will in general require modification, the integers being replaced by the elements of $\pi_2(T)$.

9. **1-Chains with coefficients from a non-abelian group.** We shall conclude with a brief account of 1-G-chains, where G is a non-abelian group, in extension of the remarks in the preceding two sections. These results were found independently by H. Whitney.

Let K be a complex as in §4. By a 1-G-chain we mean a function $f(p, q)$ defined for all ordered pairs of integers (p, q) for which $x_p x_q$ is a 1-cell of K, with values in G, and such that $f(p, q)=f^{-1}(q, p)$. We may denote such a chain by the symbol $C=\sum g_{ij} x_i x_j$, where $g_{ij}=f(i, j)$. C is a *cocycle* if running around the boundary of each 2-cell of K defines the unit element of G. The chain C is a *coboundary* if there exist h_i in G such that

$$g_{ij} = h_i^{-1} h_j \tag{9.1}$$

for all i, j such that $x_i x_j$ is a 1-cell of K. If C is a coboundary it is a cocycle. Let $C=\sum g_{ij} x_i x_j$ be a chain, and $D=\sum h_i x_i$ be a 0-G-chain; then by $C \circ D$ we mean the chain

$$\sum (h_i^{-1} g_{ij} h_j) x_i x_j. \tag{9.2}$$

If C is a cocycle, so is $C \circ D$; if C is a coboundary, so is $C \circ D$. The operation $\circ$ is associative, there is a unit, and inverses exist. If C_1 and C_2 are cocycles, and $C_1=C_2 \circ D$ for some 0-G-chain D, we say that C_1 and C_2 are *cohomologous*; in symbols, $C_1 \backsim C_2$ or $C_1/C_2=\delta D$. If G is abelian, this reduces to the ordinary notion of cohomology. The preceding remarks show that the set of 0-G-chains

form a group of transformations acting on the set of 1-G-cocycles. We shall call the class of all cocycles which are homologous to C the *coset* of C, and denote it by $[C]$. Two cosets are identical or disjoint. A chain will be called part of a cocycle if some or all of its coefficients which are 1's may be replaced by other elements of G so that the resulting chain is a cocycle.

Let f^1 and f^2 be two mappings of a complex K into an aspherical[15] space T with fundamental group G. They are homotopic if and only if the mapping F defined on $K\times 0+K\times 1$ by them may be extended throughout the 2-cells $x_ix_j\times 1$ of $K\times 1$; for since T is aspherical, it may then be extended throughout all the 3-, 4-, $\cdots$ cells. This is possible if and only if the 1-G-chain C_{f^1,f^2} with coefficients f^1_{ij} on y_iy_j and f^2_{ij} on z_iz_j is part of a cocycle. But this is so if and only if the 1-G-cocycles $C_{f^1}=\sum f^1_{ij}x_ix_j$ and $C_{f^2}=\sum f^2_{ij}x_ix_j$ of K are cohomologous. Thus we have

THEOREM 5. *The classes of mappings of K into the aspherical space T are in one-to-one correspondence with the cosets $[C]$, where the C are the 1-G-cocycles of K, and G is the fundamental group of T.*

This theorem is equivalent to the well known theorem of Brouwer and Hurewicz which states that the classes of mappings of K into T are in one-to-one correspondence with the homomorphism-classes of H = fundamental group of K into G = fundamental group of T, since the correspondence $\psi_f\leftrightarrow[C_f]$, where ψ_f denotes the class of homomorphisms derived from that induced by the mapping f under the set of inner automorphisms of G, and $C_f=\sum f_{ij}x_ix_j$, where f_{ij} is the element of G into which x_ix_j is mapped by f, is easily seen to be one-to-one. (We assume f to be normal; see §6.)

Proof. We must show that (a) if $C_f\backsim C_g$ then $\psi_f=\psi_g$, (b) if $\psi_f=\psi_g$ then $C_f\backsim C_g$.

We shall use the vertex x_0 as a fixed point in defining the fundamental group.

Ad (a): Run around any circuit in K by f, getting $f_{01}f_{12}\cdots f_{p0}$ of G. Then run around the circuit by g, getting, since $g_{ij}=h_i^{-1}f_{ij}h_j$,

$$(h_0^{-1}f_{01}h_1)(h_1^{-1}f_{12}h_2)\cdots(h_p^{-1}f_{p0}h_0) = h_0^{-1}(f_{01}f_{12}\cdots f_{p0})h_0.$$

Ad (b): Say a closed path C from x_0 in K mapped by g into the element $\beta(C)$ of G is mapped by f into $\alpha^{-1}\beta(C)\alpha$. Join each vertex x_i to x_0 by a path C_i, and let $h_i=g^{-1}(C_i)\alpha f(C_i)$. Then $f(C_i)f_{ij}f^{-1}(C_j)=\alpha^{-1}g(C_i)g_{ij}g^{-1}(C_j)\alpha$ for all i, j such that x_ix_j is a 1-cell of K. It follows that

$$f_{ij} = f^{-1}(C_i)\alpha^{-1}g(C_i)g_{ij}g^{-1}(C_j)\alpha f(C_j) = [g^{-1}(C_i)\alpha f(C_i)]^{-1}g_{ij}[g^{-1}(C_j)\alpha f(C_j)],$$

so that $C_f\backsim C_g$.

(15) See Hurewicz, loc. cit., in note 3.

NEW YORK UNIVERSITY,
NEW YORK, N. Y.

IDEALS IN BIRKHOFF LATTICES

BY

R. P. DILWORTH(1)

Introduction. In previous papers by the author (Dilworth [1, 2])(2) methods were developed for studying the arithmetical properties of Birkhoff lattices, that is, the properties of irreducibles and decompositions into irreducibles. These methods, however, required the assumption of both the ascending and descending chain conditions. In this paper we give a new technique which is applicable in general and which under the assumption of merely the ascending chain condition gives results quite as good as those of the previous work. Now the descending chain condition is equivalent to the requirement that every ideal(3) be principal. Hence if the descending chain condition does not hold we find it convenient to relate the arithmetical properties of the lattice to the structure of its lattice of ideals. Furthermore since the Birkhoff condition itself may lose much of its force if the descending chain condition does not hold, a lattice is defined to be a Birkhoff lattice if every element satisfies the Birkhoff condition(4) in the lattice of ideals. Hence if the descending chain condition holds, this definition reduces to that used in the previous papers. In the lattice of ideals, the existence of sufficient covering ideals to make the Birkhoff conditions effective can be proved.

In D1 and D2 it was shown that the arithmetical behavior of an element a was closely related to the structure of the quotient lattice $\mathfrak{S}_a$ generated by the elements covering a. Here we make a similar correlation with the structure of the quotient lattice of ideals $\mathfrak{L}_a$ generated by the ideals covering a. The important properties of $\mathfrak{S}_a$ follow from its finite dimensionality. $\mathfrak{L}_a$ on the other hand is in general *not* finite dimensional and thus one of the essential problems of the present treatment is the proof of the archimedean character of $\mathfrak{L}_a$ in the cases of interest.

If the descending chain condition holds, the Birkhoff condition is equivalent to Mac Lane's point-free exchange axiom E_5 (Mac Lane [1]). Now E_5 is independent of covering conditions, which suggests that it should be closely related to the Birkhoff condition in the lattice of ideals. We show that the Birkhoff condition in the lattice of ideals always implies E_5 and, if each principal ideal is covered by only a finite number of ideals, the two conditions are equivalent.

Presented to the Society, December 27, 1939; received by the editors May 11, 1940.

(1) Sterling Research Fellow, Yale University.

(2) These papers will be referred to as D1 and D2.

(3) An ideal is a sublattice which contains with each element all of its divisors. G. Birkhoff (Birkhoff [1]) uses the term *dual ideal* for such a sublattice.

(4) See §1, Conditions B1 and B1'.

In D1 it was shown that a lattice of finite dimensions has unique irreducible decompositions if and only if it is a Birkhoff lattice in which every modular sublattice is distributive. This result no longer holds if we drop the descending chain condition as we show by an example. However, by strengthening slightly the condition that every modular sublattice be distributive, we have the following theorem:

THEOREM 6.6. *Let $\mathfrak{S}$ satisfy the ascending chain condition. Then every element of $\mathfrak{S}$ is uniquely expressible as a reduced crosscut of irreducibles if and only if the following conditions hold.*

E_5. (Mac Lane's point-free exchange axiom.) $a\supset b\supset a\cap c$, $c\neq a\cap c$ *implies that* $c_1\neq a\cap c$ *exists such that* $c\supset c_1\supset a\cap c$ *and* $b=a\cap(b\cup c_1)$.

A. $a\cup b\supset x\supset a\cap b$, $a\cap x=b\cap x=a\cap b$ *implies* $x=a\cap b$.

If we go over to the lattice of ideals, E_5 may be replaced by the condition that $\mathfrak{S}$ be a Birkhoff lattice, and A, by the requirement that the ideals covering a principal ideal generate a Boolean algebra.

In D2, Birkhoff lattices in which the number of components in the irreducible decompositions of each element is unique were characterized in terms of the structure of the quotient lattices $\mathfrak{S}_a$. We prove here:

THEOREM 5.1. *Let $\mathfrak{S}$ be a Birkhoff lattice satisfying the ascending chain condition and let $\mathfrak{L}$ denote its lattice of ideals. Then the number of components in the irreducible decompositions of each element of the lattice $\mathfrak{S}$ is unique if and only if the ideals covering any principal ideal of the lattice $\mathfrak{L}$ generate a dense, modular sublattice of $\mathfrak{L}$.*

By means of ideal methods we give a new proof of the Kurosch-Ore decomposition theorem for modular lattices in its most general form. The proof rests on the fact that if an element of a modular lattice has a decomposition into irreducibles then the sublattice generated by the ideals covering the element is of finite dimensions.

Finally §§7 and 8 contain examples which show the complications which may arise when the descending chain condition does not hold.

1. **Notation and definitions.** The fixed lattice of elements $a, b, c, \cdots$ will be denoted by $\mathfrak{S}$. $\cup$ and $\cap$ will denote union and cross-cut in place of the symbols (,) and [,] used in D1 and D2. $\supset$ denotes lattice division. $a=b$ is defined by the two formulas $a\supset b$, $b\supset a$. If $a\supset b$, $a\neq b$ and $a\supset x\supset b$ implies $a=x$ or $x=b$, we say that *a covers b* and write $a>b$. Elements which cover the null element z of a lattice are called *points* and elements covered by the unit element u are said to be *simple*.

A lattice $\mathfrak{S}$ satisfies the ascending (descending) chain condition if every chain $a_1\subset a_2\subset a_3\subset \cdots$ $(a_1\supset a_2\supset a_3\supset \cdots)$ has only a finite number of distinct elements. If both the ascending and descending chain conditions hold, $\mathfrak{S}$ is said to be *archimedean* or of *finite dimensions*.

Throughout the paper we shall be particularly interested in lattices which satisfy the following weak form of the modular axiom.

B1. $a > a \cap b \rightarrow a \cup b > b$(5).

Another form of B1 is the following:

B1′. $b > a,\ c \supset a,\ c \not\supset b \rightarrow b \cup c > c$.

If B1′ is satisfied for a given a and any b and c we say that a satisfies the *Birkhoff condition* in $\mathfrak{S}$. Hence B1 holds in $\mathfrak{S}$ if and only if each element of $\mathfrak{S}$ satisfies the Birkhoff condition.

We state now some lemmas on elements satisfying the Birkhoff condition which are refinements of Lemmas 3.1–3.3 of D2.

LEMMA 1.1. *Let a satisfy the Birhoff condition in $\mathfrak{S}$ and let $a_1, \cdots, a_k > a$. Then each union independent*(6) *set of the a_i is contained in a maximal independent set.*

The usual proof is valid under the weaker hypotheses of the lemma.

LEMMA 1.2. *Let a satisfy the Birkhoff condition and let $a_1, \cdots, a_k > a$. Then each union independent set of the a_i generates a Boolean algebra.*

We note that the usual proof (for example Theorem 2.3 of D1) is not valid in this case since it depends upon the existence of a rank function. Under the hypotheses of the lemma, complete chains need not have the same length and hence a rank function will in general not exist.

Now let A and B be two arbitrary subsets of the set $\{a_1, \cdots, a_k\}$. Let $\Sigma(A)$ denote the union of the elements of A and denote the set-theoretic union and cross-cut of A and B by $A \cup B$ and $A \cap B$ respectively. We shall show that

$$\Sigma(A) \cap \Sigma(B) = \Sigma(A \cap B). \tag{1}$$

Let $\mu(A)$ denote the number of elements in A and set $\nu(A) = k - \mu(A)$. If $\nu(A \cap B) = 0$, then $\mu(A \cap B) = k$ and $A = B$. Hence (1) holds. If $\nu(A \cap B) = 1$, then either $A \supset B$ or $B \supset A$ and again (1) holds. Now let (1) hold for all A and B such that $\nu(A \cap B) < l$. Let $\nu(A \cap B) = l$ for some A and B. Then $\mu(A \cap B) = k - l = r$. Hence $A = \{a_1, \cdots, a_r, a_{r+1}, \cdots, a_s\}$ and $B = \{a_1, \cdots, a_r, a'_{r+1}, \cdots, a'_t\}$. Since (1) is trivial if $B \supset A$, we may assume that $s > r$. Let $B' = \{a_1, \cdots, a_r, a_{r+1}, a'_{r+1}, \cdots, a'_t\}$. Now $\mu(A \cap B') = r+1$ and hence $\nu(A \cap B') = k - (r+1) = l - 1 < l$. By the induction assumption $\Sigma(A \cap B') = \Sigma(A) \cap \Sigma(B')$. Thus $\Sigma(A \cap B') = \Sigma(A) \cap \Sigma(B') \supset \Sigma(A) \cap \Sigma(B) \supset \Sigma(A \cap B)$.

---

(5) $\rightarrow$ denotes formal implication.

(6) A set of elements $x_1, \cdots, x_n$ is said to be ***union independent*** or simply ***independent*** if $x_1 \cup \cdots \cup x_{i-1} \cup x_{i+1} \cup \cdots \cup x_n \not\supset x_i,\ i = 1, \cdots, n$. Similarly the set is said to be ***cross-cut independent*** if $x_i \not\supset x_1 \cap \cdots \cap x_{i-1} \cap x_{i+1} \cap \cdots \cap x_n,\ i = 1, \cdots, n$.

Since $a_1, \cdots, a_k$ are independent we have $\Sigma(A \cap B) \not\supset a_{r+1}$ and hence $\Sigma(A \cap B') = a_{r+1} \cup \Sigma(A \cap B) > \Sigma(A \cap B)$. If $\Sigma(A \cap B') = \Sigma(A) \cap \Sigma(B)$, then $\Sigma(B) \supset a_{r+1}$ contrary to the independence of $a_1, \cdots, a_k$. Hence $\Sigma(A) \cap \Sigma(B) = \Sigma(A \cap B)$. Thus (1) holds for $\nu(A \cap B) = l$ and by induction (1) holds for all A and B. Clearly $\Sigma(A) \cup \Sigma(B) = \Sigma(A \cup B)$. If $\Sigma(A) = \Sigma(B)$, then $A = B$ by the independence of $a_1, \cdots, a_k$. Hence the elements which can be expressed as a union of the a_i are isomorphic to the subsets of $a_1, \cdots, a_k$ under union and cross-cut and thus $a_1, \cdots, a_k$ generate a Boolean algebra. This completes the proof of the lemma.

LEMMA 1.3. *Let a satisfy the Birkhoff condition and let $a_1, \cdots, a_k > a$. Then any two maximal union independent sets of the a_i have the same number of elements and any element of one set may be replaced by a suitably chosen element of the other without altering the maximal property.*

The usual proof is valid in this case.

LEMMA 1.4. *Let a satisfy the Birkhoff condition and let $a_1, \cdots, a_k > a$. Then any chain joining $a_1 \cup \cdots \cup a_k$ to a has not more than $k+1$ distinct members.*

We may clearly suppose that $a_1, \cdots, a_k$ are independent. Let $a = b_0 \subset b_1 \subset b_2 \subset \cdots \subset b_{l-1} \subset b_l = a_1 \cup \cdots \cup a_k$ be a chain joining $a_1 \cup \cdots \cup a_k$ to a having $l+1$ distinct members and let us assume that $l > k$. Clearly $b_0 < b_0 \cup a_1 < \cdots < b_0 \cup a_1 \cup \cdots \cup a_{k-1} < a_1 \cup \cdots \cup a_k$ by the Birkhoff condition. Now suppose that it has been shown that $a \subset b_1 \subset \cdots \subset b_i < b_i \cup a_1 < \cdots < b_i \cup a_1 \cup \cdots \cup a_{k_i-1} < a_1 \cup \cdots \cup a_k$ where $k_i \leqq k-i$ and $i < k$. Consider the chain $a \subset b_1 \subset \cdots \subset b_i \subset b_{i+1} \subset b_{i+1} \cup a_1 \subset \cdots \subset b_{i+1} \cup a_1 \cup \cdots \cup a_{k_i-1} \subset a_1 \cup \cdots \cup a_k$. Let us assume that all of the members of this chain are distinct. If $b_i \cup a_1 \cup \cdots \cup a_{k_i-1} \not\supset b_{i+1}$, then $a_1 \cup \cdots \cup a_k \supset e_{i+1} \cup a_1 \cup \cdots \cup a_{k_i-1} \supset b_i \cup a_1 \cup \cdots \cup a_{k_i-1}$ and $b_{i+1} \cup a_1 \cup \cdots \cup a_{k_i-1} \neq b_i \cup a_i \cup \cdots \cup a_{k_i-1}$. But $a_1 \cup \cdots \cup a_k > b_i \cup a_1 \cup \cdots \cup a_{k_i-1}$ and hence $a_1 \cup \cdots \cup a_k = b_{i+1} \cup a_1 \cup \cdots \cup a_{k_i-1}$ contrary to our assumption. Thus $b_i \cup a_1 \cup \cdots \cup a_{k_i-1} \supset b_{i+1}$. If $b_i \cup a_1 \cup \cdots \cup a_{k_i-2} \not\supset b_{i+1}$, we have $b_i \cup a_1 \cup \cdots \cup a_{k_i-1} \supset b_{i+1} \cup a_1 \cup \cdots \cup a_{k_i-2} \supset b_i \cup a_1 \cup \cdots \cup a_{k_i-2}$. But $b_i \cup a_1 \cup \cdots \cup a_{k_i-1} > b_i \cup a_1 \cup \cdots \cup a_{k_i-2}$ and hence $b_i \cup a_1 \cup \cdots \cup a_{k_i-1} = b_{i+1} \cup a_1 \cup \cdots \cup a_{k_i-2}$ contrary to our assumption. Thus $b_i \cup a_1 \cup \cdots \cup a_{k_i-2} \supset b_{i+1}$. Continuing in this manner we eventually have $b_i \cup a_1 \supset b_{i+1}$. But then $b_i \cup a_1 \supset b_{i+1} \supset b_i$ and $b_{i+1} \neq b_i$. Hence $b_{i+1} = b_i \cup a_1 = b_{i+1} \cup a_1$ which contradicts our assumption. We conclude, then, that at least two members of the above chain are equal. Thus (renumbering the a's if necessary) using the Birkhoff condition we have $a \subset b_1 \subset \cdots \subset b_i \subset b_{i+1} < b_{i+1} \cup a_1 < \cdots < b_{i+1} \cup a_1 \cup \cdots \cup a_{k_{i+1}-1} < a_1 \cup \cdots \cup a_k$ where $k_{i+1} \leqq k_i - 1 \leqq k - (i+1)$. By induction, we get $a = b_0 \subset b_1 \subset \cdots \subset b_{r-1} < a_1 \cup \cdots \cup a_k$ where $r \leqq k$. But then $b_r = a_1 \cup \cdots \cup a_k$ and hence $r = l$ which contradicts $l > k$. Thus $l \leqq k$ and the lemma follows.

The dual of condition B1 is the condition

B2. $a \cup b > b \rightarrow a > a \cap b$.

G. Birkhoff (Birkhoff [2]) has proved the following lemma which relates B1 and B2 to modularity.

LEMMA 1.5. *An archimedean lattice $\mathfrak{S}$ is modular if and only if* B1 *and* B2 *are satisfied.*

2. Lattice ideals. A sublattice $\mathfrak{a}$ of $\mathfrak{S}$ is said to be an *ideal* if $x \supset a$, $a \in \mathfrak{a}$ implies $x \in \mathfrak{a}$. If $\mathfrak{a}$ consists of all elements x such that $x \supset a$ for a fixed a, then $\mathfrak{a}$ is said to be a *principal* ideal and we write $\mathfrak{a} = (a)$. Now suppose that $\mathfrak{S}$ satisfies the descending chain condition. Then the set of elements in $\mathfrak{a}$ has a cross-cut which can be expressed as a cross-cut of a finite number of them and hence belongs to $\mathfrak{a}$. Thus $\mathfrak{a}$ consists of all divisors of a fixed element of $\mathfrak{S}$ and hence is principal. Conversely, if every ideal of $\mathfrak{S}$ is principal, then a descending chain $a_1 \supset a_2 \supset \cdots$ generates an ideal $\mathfrak{a}$ which consists of all x such that $x \supset a_k$ for some k. But then $\mathfrak{a} = (a)$ and $a \supset a_k$ for some k. Hence $a = a_k = a_{k+1} = \cdots$ and every descending chain has only a finite number of distinct elements. We thus have

LEMMA 2.1. *$\mathfrak{S}$ satisfies the descending chain condition if and only if every ideal is principal.*

The set of ideals of $\mathfrak{S}$ will be denoted by $\mathfrak{L}$.

DEFINITION 2.1. *The union $\mathfrak{a} \cup \mathfrak{b}$ of two ideals $\mathfrak{a}$ and $\mathfrak{b}$ is the set of all elements x such that $x \supset a \cup b$ for some $a \in \mathfrak{a}$ and $b \in \mathfrak{b}$. Similarly the cross-cut $\mathfrak{a} \cap \mathfrak{b}$ is the set of all elements y such that $y \supset a \cap b$ for some $a \in \mathfrak{a}$ and $b \in \mathfrak{b}$.*

It is readily verified that the union and cross-cut so defined are ideals and that $\mathfrak{L}$ is a lattice under these operations. The union $\mathfrak{a} \cup \mathfrak{b}$ is simply the set-theoretic cross-cut of $\mathfrak{a}$ and $\mathfrak{b}$.

The definition of cross-cut may be readily extended to any subset S of $\mathfrak{L}$. $\Pi(S)$ consists of all elements of $\mathfrak{S}$ which belong to the cross-cut of a finite number of ideals of S. If $\mathfrak{S}$ has a unit element u, the union $\Sigma(S)$ is also defined and is simply the set-theoretic cross-cut of the ideals of S.

If $\mathfrak{a}$ and $\mathfrak{b}$ are principal ideals $\mathfrak{a} = (a)$ and $\mathfrak{b} = (b)$, then by Definition 2.1 $\mathfrak{a} \cup \mathfrak{b} = (a \cup b)$ and $\mathfrak{a} \cap \mathfrak{b} = (a \cap b)$. Hence the set of principal ideals forms a sublattice of $\mathfrak{L}$ which is isomorphic to $\mathfrak{S}$ and we may thus consider $\mathfrak{S}$ as a sublattice of $\mathfrak{L}$.

LEMMA 2.2. *$\mathfrak{L}$ is a modular (distributive) if and only if $\mathfrak{S}$ is modular (distributive).*

Since $\mathfrak{S}$ is a sublattice of $\mathfrak{L}$, the modularity (distributivity) of $\mathfrak{L}$ implies the modularity (distributivity) of $\mathfrak{S}$.

Now let $\mathfrak{S}$ be distributive and let $x \in \mathfrak{a} \cup (\mathfrak{b} \cap \mathfrak{c})$. Then $x \supset a \cup (b \cap c)$ where $a \in \mathfrak{a}$, $b \in \mathfrak{b}$ and $c \in \mathfrak{c}$ by Definition 2.1. But then $x \supset a \cup (b \cap c) \supset (a \cup b) \cap (a \cup c)$ since $\mathfrak{S}$ is distributive and hence $x \in (\mathfrak{a} \cup \mathfrak{b}) \cap (\mathfrak{a} \cup \mathfrak{c})$. Thus $\mathfrak{a} \cup (\mathfrak{b} \cap \mathfrak{c}) \supset (\mathfrak{a} \cup \mathfrak{b}) \cap (\mathfrak{a} \cup \mathfrak{c})$. But $(\mathfrak{a} \cup \mathfrak{b}) \cap (\mathfrak{a} \cup \mathfrak{c}) \supset \mathfrak{a} \cup (\mathfrak{b} \cap \mathfrak{c})$ trivially. Hence $\mathfrak{L}$ is distributive. Now let $\mathfrak{S}$ be modular. Suppose $\mathfrak{a} \supset \mathfrak{b}$ and $x \in \mathfrak{b} \cup (\mathfrak{a} \cap \mathfrak{c})$. Then $x \supset b \cup (a \cap c)$ where $a \in \mathfrak{a}$, $b \in \mathfrak{b}$ and $c \in \mathfrak{c}$ by Definition 2.1. Now since $\mathfrak{a} \supset \mathfrak{b}$ we have $a \supset b_1$ where $b_1 \in \mathfrak{b}$. But then $x \supset (b \cap b_1) \cup (a \cap c)$ and $a \supset b \cap b_1$ where $b_1 \cap b \in \mathfrak{b}$. Hence $x \supset a \cap ((b \cap b_1) \cup c)$ since $\mathfrak{S}$ is modular and $x \in \mathfrak{a} \cap (\mathfrak{b} \cup \mathfrak{c})$. Thus $\mathfrak{b} \cup (\mathfrak{a} \cap \mathfrak{c}) \supset \mathfrak{a} \cap (\mathfrak{b} \cup \mathfrak{c})$ and since $\mathfrak{a} \cap (\mathfrak{b} \cup \mathfrak{c}) \supset \mathfrak{b} \cup (\mathfrak{a} \cap \mathfrak{c})$ trivially, $\mathfrak{L}$ is modular. This completes the proof.

LEMMA 2.3. *Let* $\mathfrak{a} \supset \mathfrak{b} \supset \cdots \supset \mathfrak{u} \supset \cdots$ *be a chain of ideals such that* $\mathfrak{u} \supset (a)$ *and* $\mathfrak{u} \neq (a)$ *for all ideals of the chain. Then if* $\mathfrak{p}$ *is the cross-cut of the ideals of the chain,* $\mathfrak{p} \supset (a)$ *and* $\mathfrak{p} \neq (a)$.

We note that $\mathfrak{p}$ is the set-theoretic union of the elements of the ideals $\mathfrak{a}, \mathfrak{b}, \cdots, \mathfrak{u}, \cdots$. For if $x \in \mathfrak{p}$, then x divides a finite cross-cut of the ideals of the chain and hence divides some ideal of the chain. Now suppose $\mathfrak{p} = a$. Then $a \in \mathfrak{p}$ and $a \in \mathfrak{u}$ for some $\mathfrak{u}$. But then $\mathfrak{u} = (a)$ contrary to assumption. Hence $\mathfrak{p} \neq a$.

The results so far have been independent of the well ordering hypothesis. However, to prove the fundamental property of the ideals we must assume that the elements of $\mathfrak{S}$ can be well ordered. This will be assumed through the remainder of the paper.

THEOREM 2.1. *Let* $\mathfrak{b} \supset (a)$ *and* $\mathfrak{b} \neq (a)$. *Then there exists an ideal* $\mathfrak{p}$ *such that* $\mathfrak{b} \supset \mathfrak{p} > (a)$.

Proof. Let U be the set of all elements x such that $x \supset a$. Let U be well ordered, $U = \{x_\nu\}$, $\nu < \sigma$. Define $\mathfrak{a}_0 = \mathfrak{b}$. Now suppose that $\mathfrak{a}_\mu$ has been defined for all $\mu < \nu$ in such a way that $\mathfrak{a}_\mu \neq (a)$, $\mathfrak{a}_\mu \supset \mathfrak{a}_{\mu'}$ if $\mu \leqq \mu'$, and $\mathfrak{a}_\mu \cap x_\mu = \mathfrak{a}_\mu$ or $\mathfrak{a}_\mu \cap x_\mu = a$. Let $\mathfrak{c}_\nu$ be the cross-cut of all $\mathfrak{a}_\mu$ with $\mu < \nu$. Then $\mathfrak{c}_\nu \neq (a)$ by Lemma 2.3. If $\mathfrak{c}_\nu \cap x_\nu \neq (a)$, let $\mathfrak{a}_\nu = \mathfrak{c}_\nu \cap x_\nu$; otherwise let $\mathfrak{a}_\nu = \mathfrak{c}_\nu$. Then $\mathfrak{a}_\nu \neq (a)$ and $\mathfrak{a}_\mu \supset \mathfrak{a}_\nu$, all $\mu < \nu$. Clearly $\mathfrak{a}_\nu \cap x_\nu = a$ or $\mathfrak{a}_\nu$. Now let $\mathfrak{p} = \Pi_{\nu < \sigma} \mathfrak{a}_\nu$. Then $\mathfrak{p} \neq a$ by Lemma 2.3 and $\mathfrak{b} \supset \mathfrak{p}$. If $\mathfrak{p} \supset \mathfrak{a} \supset (a)$ and $\mathfrak{p} \neq \mathfrak{a}$, there exists an element $x \in \mathfrak{a}$ such that $x \notin \mathfrak{p}$. Since $x \supset a$ we have $x = x_\nu$ for some ν. But then $\mathfrak{a}_\nu \cap x = a$ since otherwise $x_\nu \supset \mathfrak{a}_\nu \supset \mathfrak{p}$ which contradicts $x \notin \mathfrak{p}$. Thus $a = \mathfrak{a}_\nu \cap x \supset \mathfrak{p} \cap \mathfrak{a} = \mathfrak{a} \supset a$ and $\mathfrak{a} = (a)$. Hence $\mathfrak{p} > a$.

In the special instances of Boolean algebras and distributive lattices, Theorem 2.1 gives respectively the existence of the prime ideals of Stone (Stone [1]) and the maximal collections of Wallman (Wallman [1]).

We next prove a theorem which enables us to pass from ideal relations to the corresponding element relations. The following lemma is required.

LEMMA 2.4. *Let* $\mathfrak{a} = \mathfrak{a}(\mathfrak{a}_1, \cdots, \mathfrak{a}_n)$ *be an ideal obtained from the ideals*

$\mathfrak{a}_1, \cdots, \mathfrak{a}_n$ *by forming a finite number of unions and cross-cuts. Then if* $x \in \mathfrak{a}$, *there exist elements* $a_1, \cdots, a_n, a_i \in \mathfrak{a}_i$, *such that* $x \supset \mathfrak{a}(a_1, \cdots, a_n)$.

For let $n(\mathfrak{a})$ denote the number of union and cross-cut symbols in the expression $\mathfrak{a}(\mathfrak{a}_1, \cdots, \mathfrak{a}_n)$. Suppose that the lemma is true for all expressions $\mathfrak{a}$ for which $n(\mathfrak{a}) < k$. Let $n(\mathfrak{a}) = k$. Then $\mathfrak{a} = \mathfrak{a}_1 \circ \mathfrak{a}_2$ where $\circ$ is either $\cap$ or $\cup$ and $n(\mathfrak{a}_1) < k$, $n(\mathfrak{a}_2) < k$. Now if $x \in \mathfrak{a}$ we have $x \supset x_1 \circ x_2$ where $x_1 \in \mathfrak{a}_1$ and $x_2 \in \mathfrak{a}_2$ by the definition of union and cross-cut. But then by the induction assumption elements $a_1', \cdots, a_n'$ and $a_1'', \cdots, a_n''$ exist such that $x_1 \supset \mathfrak{a}_1(a_1', \cdots, a_n')$, $x_2 \supset \mathfrak{a}_2(a_1'', \cdots, a_n'')$. Let $a_i = a_i' \cap a_i''$. Then $x \supset x_1 \circ x_2 \supset \mathfrak{a}_1(a_1', \cdots, a_n') \circ \mathfrak{a}_2(a_1'', \cdots, a_n'') \supset \mathfrak{a}_1(a_1, \cdots, a_n) \circ \mathfrak{a}_2(a_1, \cdots, a_n) = \mathfrak{a}(a_1, \cdots, a_n)$ and a_i is clearly in $\mathfrak{a}_i$. Since the lemma is trivially true when $n(\mathfrak{a}) = 1$ by Definition 2.1, the proof is complete.

THEOREM 2.2. *Let* $(a) = \mathfrak{a}(\mathfrak{a}_1, \cdots, \mathfrak{a}_n)$ *where* $\mathfrak{a}$ *is obtained from* $\mathfrak{a}_1, \cdots, \mathfrak{a}_n$ *by forming a finite number of union and cross-cuts. Then* $(a) = \mathfrak{a}(a_1, \cdots, a_n)$ *where* $a_i \in \mathfrak{a}_i$.

Proof. By Lemma 2.4 $a \supset \mathfrak{a}(a_1, \cdots, a_n)$ where $a_i \in \mathfrak{a}_i$. But then $\mathfrak{a}(a_1, \cdots, a_n) \supset \mathfrak{a}(\mathfrak{a}_1, \cdots, \mathfrak{a}_n) = (a)$. Hence $(a) = \mathfrak{a}(a_1, \cdots, a_n)$.

As an example, if $a = \mathfrak{a}_1 \cap \cdots \cap \mathfrak{a}_n$ then elements $a_i \in \mathfrak{a}_i$ exist such that $a = a_1 \cap \cdots \cap a_n$.

We conclude this section with two useful lemmas on irreducibles(7).

LEMMA 2.5. *If* q *is irreducible in* $\mathfrak{S}$, *then* q *is irreducible in* $\mathfrak{L}$.

For if q is reducible in $\mathfrak{L}$, then $q = \mathfrak{a} \cap \mathfrak{b}$, $\mathfrak{a}$, $\mathfrak{b} \neq q$. But then $q = a \cap b$, $a \in \mathfrak{a}$, $b \in \mathfrak{b}$ by Theorem 2.2. Clearly $a \neq q$ and $b \neq q$. Hence q is reducible in $\mathfrak{S}$. Inverting the logic gives the lemma.

LEMMA 2.6. *Let every element of* $\mathfrak{S}$ *be expressible as a cross-cut of irreducibles. Then if* $\mathfrak{a} \supset \mathfrak{b}$, $\mathfrak{a} \neq \mathfrak{b}$, *there exists an irreducible* q *of* $\mathfrak{S}$ *such that* $q \supset \mathfrak{b}$, $q \not\supset \mathfrak{a}$.

For since $\mathfrak{a} \neq \mathfrak{b}$, b exists such that $b \in \mathfrak{b}$, $b \notin \mathfrak{a}$. Let $b = q_1 \cap \cdots \cap q_k$. If $q_i \in \mathfrak{a}$ for every i then $b \in \mathfrak{a}$ contrary to assumption. Hence $q_i \notin \mathfrak{a}$ for some i. But then $q_i \supset b \supset \mathfrak{b}$.

3. **Birkhoff lattices.** In D1 and D2 a lattice satisfying B1 was defined to be a Birkhoff lattice. Since both the ascending and descending chain conditions were assumed to hold, B1 was never satisfied trivially. Now in a sufficiently general lattice no covering relations may exist and B1 will hold vacuously. Hence we formulate a more general definition which reduces to that used in D1 and D2 if the descending chain condition holds.

DEFINITION 3.1. *A lattice* $\mathfrak{S}$ *is said to be a Birkhoff lattice if each element of* $\mathfrak{S}$ *satisfies the Birkhoff condition in the lattice of ideals.*

(7) An element q is said to be *cross-cut irreducible* or simply *irreducible* if $q = a \cap b \rightarrow q = a$ or $q = b$. q is said to be *union irreducible* if $q = a \cup b \rightarrow q = a$ or $q = b$.

A lattice $\mathfrak{S}$ is never vacuously a Birkhoff lattice since by Theorem 2.1 covering ideals always exist. Furthermore if the descending chain condition holds, then every ideal is principal and $\mathfrak{S}$ is a Birkhoff lattice if and only if B1 holds in $\mathfrak{S}$.

Now if $\mathfrak{S}$ has a unit element u and a is any element of $\mathfrak{S}$, then the union of the ideals covering a exists and will be denoted by $\mathfrak{u}_a$. Let $\mathfrak{L}_a$ denote the quotient lattice of all ideals of $\mathfrak{L}$ which are divisible by $\mathfrak{u}_a$ and which divide a. Then $\mathfrak{L}_a$ is a dense sublattice of $\mathfrak{L}$ and every proper divisor of a in $\mathfrak{L}$ divides some point ideal of $\mathfrak{L}_a$ by Theorem 2.1. Clearly $\mathfrak{L}_a$ reduces to the sublattice $\mathfrak{S}_a$ of the previous papers if the descending chain condition holds. The essential properties of $\mathfrak{S}_a$ followed from its finite dimensionality. But $\mathfrak{L}_a$ is in general *not* finite dimensional. However we now prove a theorem which insures the archimedean character of $\mathfrak{L}_a$ in most cases of arithmetical interest. We need the following lemma:

LEMMA 3.1. *Let $\mathfrak{S}$ be a Birkhoff lattice. Then if $\mathfrak{p}_1, \cdots, \mathfrak{p}_k$ is a maximal independent set of point ideals of $\mathfrak{L}_a$, the length of any chain of $\mathfrak{L}_a$ is not greater than k.*

Since the length of any chain is one less than the number of distinct members of the chain, the lemma follows immediately from Lemma 1.4 and Definition 3.1.

According to Lemma 3.1, $\mathfrak{L}_a$ is archimedean if and only if $\mathfrak{u}_a$ can be expressed as a union of a finite number of point ideals of $\mathfrak{L}_a$.

THEOREM 3.1. *Let $\mathfrak{S}$ be a Birkhoff lattice in which every element may be represented as a cross-cut of irreducibles. Then $\mathfrak{L}_a$ is archimedean if and only if the number of components in the irreducible decompositions of a is bounded.*

Proof. Let the number of components in the irreducible decompositions of a be bounded, say less than n. Then if $\mathfrak{L}_a$ is not archimedean, by Lemmas 1.2 and 3.1 there are n union independent point ideals $\mathfrak{p}_1, \cdots, \mathfrak{p}_n$ of $\mathfrak{L}_a$ which generate a Boolean algebra. Let $\mathfrak{a}_i = \mathfrak{p}_1 \cup \cdots \cup \mathfrak{p}_{i-1} \cup \mathfrak{p}_{i+1} \cup \cdots \cup \mathfrak{p}_n$. Then $a = \mathfrak{a}_1 \cap \cdots \cap \mathfrak{a}_n$. Hence by Theorem 2.2 $a = a_1 \cap \cdots \cap a_n$ where $a_i \in \mathfrak{a}_i$. Now let $a_i = q_{i1} \cap \cdots \cap q_{ik_i}$ where $q_{i1}, \cdots, q_{ik_i}$ are irreducibles of $\mathfrak{S}$. Then $a = q_{11} \cap q_{12} \cap \cdots \cap q_{nk_n}$ and this representation may be reduced([8]) by dropping our superfluous irreducibles. However not all of the irreducibles belonging to any one a_i may be dropped out since otherwise $a = q_{11} \cap \cdots \cap q_{nk_n} \supset a_1 \cap \cdots \cap a_{i-1} \cap a_{i+1} \cap \cdots \cap a_n \supset \mathfrak{a}_1 \cap \cdots \cap \mathfrak{a}_{i-1} \cap \mathfrak{a}_{i+1} \cap \cdots \cap \mathfrak{a}_n \supset \mathfrak{p}_i$ contrary to $\mathfrak{p}_i > a$. Hence a has a decomposition having at least n components. But this contradicts our assumption that the number of components is less than n. Hence $\mathfrak{L}_a$ is archimedean and of length less than n.

On the other hand let the number of components be unbounded. Then for

([8]) A representation $a = a_1 \cap a_2 \cap \cdots \cap a_n$ is said to be reduced if $a_1, \cdots, a_n$ are cross-cut independent.

every k there is an irreducible decomposition $a=q_1\cap\cdots\cap q_n$ with $n\geqq k$. Let $q_i'=q_1\cap\cdots\cap q_{i-1}\cap q_{i+1}\cap\cdots\cap q_n$. Then $q_i'\supset a$ and $q_i'\neq a$ since the representation is reduced. Hence $q_i'\supset\mathfrak{p}_i>a$ by Theorem 2.1. Suppose $\mathfrak{p}_1\cup\cdots\cup\mathfrak{p}_{i-1}\cup\mathfrak{p}_{i+1}\cup\cdots\cup\mathfrak{p}_n\supset\mathfrak{p}_i$. Then $q_i\supset q_1'\cup\cdots\cup q_{i-1}'\cup q_{i+1}'\cup\cdots\cup q_n'\supset\mathfrak{p}_1\cup\cdots\cup\mathfrak{p}_{i-1}\cup\mathfrak{p}_{i+1}\cup\cdots\cup\mathfrak{p}_n\supset\mathfrak{p}_i$ and $a=q_i\cap q_i'\supset\mathfrak{p}_i$ which contradicts $\mathfrak{p}_i>a$. Thus $\mathfrak{p}_1,\cdots,\mathfrak{p}_n$ are union independent. Hence for every k there are more than k union independent point ideals of $\mathfrak{L}_a$ and $\mathfrak{L}_a$ is *not* archimedean.

If $\mathfrak{L}_a$ is archimedean it has some simple structure properties which follow from the Birkhoff condition.

THEOREM 3.2. *Let $\mathfrak{S}$ be a Birkhoff lattice. Then if $\mathfrak{L}_a$ is archimedean, it is complemented and every ideal can be expressed as a cross-cut of simple ideals.*

Proof. Let $\mathfrak{a}\in\mathfrak{L}_a$ and let $\mathfrak{p}_1,\cdots,\mathfrak{p}_k$ be a maximal independent set of point ideals of $\mathfrak{L}_a$ divisible by $\mathfrak{a}$. Imbed $\mathfrak{p}_1,\cdots,\mathfrak{p}_k$ in a maximal independent set $\mathfrak{p}_1,\cdots,\mathfrak{p}_n$. Let $\mathfrak{a}'=\mathfrak{p}_{k+1}\cup\cdots\cup\mathfrak{p}_n$. Then $\mathfrak{a}\cup\mathfrak{a}'\supset\mathfrak{p}_1\cup\cdots\cup\mathfrak{p}_n\supset\mathfrak{u}_a$. Hence $\mathfrak{a}\cup\mathfrak{a}'=\mathfrak{u}_a$. Now suppose that $\mathfrak{a}\cap\mathfrak{a}'\neq a$. Then $\mathfrak{a}\cap\mathfrak{a}'\supset\mathfrak{p}>a$ by Theorem 2.1. Since $\mathfrak{a}\supset\mathfrak{p}$ we have $\mathfrak{p}_1\cup\cdots\cup\mathfrak{p}_k\supset\mathfrak{p}$ by the maximal property of $\mathfrak{p}_1,\cdots,\mathfrak{p}_k$ and $a=(\mathfrak{p}_1\cup\cdots\cup\mathfrak{p}_k)\cap\mathfrak{a}'\supset\mathfrak{p}$, which contradicts $\mathfrak{p}>a$. Hence $\mathfrak{a}\cap\mathfrak{a}'=a$ and $\mathfrak{L}_a$ is complemented.

Now let $\mathfrak{q}$ be irreducible in $\mathfrak{L}_a$. Let $\mathfrak{p}_1,\cdots,\mathfrak{p}_k$ be a maximal independent set of point ideals of $\mathfrak{L}_a$ divisible by $\mathfrak{q}$ and let this set be imbedded in a maximal independent set $\mathfrak{p}_1,\cdots,\mathfrak{p}_k,\cdots,\mathfrak{p}_n$. Then $\mathfrak{q}\not\supset\mathfrak{p}_{k+1},\cdots,\mathfrak{p}_n$ and hence $\mathfrak{q}\cup\mathfrak{p}_i>\mathfrak{q}$, $i=k+1,\cdots,n$, by B1′. But since $\mathfrak{q}$ is irreducible in $\mathfrak{L}_a$ we have $\mathfrak{q}\cup\mathfrak{p}_{k+1}=\cdots=\mathfrak{q}\cup\mathfrak{p}_n$. Hence $\mathfrak{u}_a=\mathfrak{q}\cup\mathfrak{u}_a=\mathfrak{q}\cup\mathfrak{p}_{k+1}\cup\cdots\cup\mathfrak{q}\cup\mathfrak{p}_n=\mathfrak{q}\cup\mathfrak{p}_{k+1}>\mathfrak{q}$. Thus each ideal which is irreducible in $\mathfrak{L}_a$ is a simple ideal of $\mathfrak{L}_a$ and since $\mathfrak{L}_a$ is archimedean each ideal of $\mathfrak{L}_a$ can be represented as a cross-cut of simple ideals.

If $\mathfrak{L}_a$ is not archimedean it will in general neither be complemented nor will every ideal be expressible as a cross-cut of simple ideals[9]. In the archimedean case an arbitrary complement of $\mathfrak{a}$ in $\mathfrak{L}_a$ will be denoted by $\mathfrak{a}'$.

DEFINITION 3.2. *An ideal $\mathfrak{c}\neq\mathfrak{u}_a$ of $\mathfrak{L}_a$ is said to be characteristic if there exists an irreducible q of $\mathfrak{S}$ which divides exactly the same point ideals of $\mathfrak{L}_a$ as $\mathfrak{c}$.*

THEOREM 3.3. *An element $a\in\mathfrak{S}$ has a reduced representation $a=q_1\cap\cdots\cap q_n$ where $q_1,\cdots,q_n$ are irreducibles if and only if a has a reduced representation $a=\mathfrak{c}_1\cap\cdots\cap\mathfrak{c}_n$ where $\mathfrak{c}_1,\cdots,\mathfrak{c}_n$ are characteristic ideals of $\mathfrak{L}_a$ such that $q_i\supset\mathfrak{c}_i$.*

Proof. Let $a=q_1\cap\cdots\cap q_n$ be a reduced representation of a as a cross-cut of irreducibles. If $q_i\supset\mathfrak{u}_a$ for some i, then $q_1\cap\cdots\cap q_{i-1}\cap q_{i+1}\cap\cdots\cap q_n\supset\mathfrak{p}_i>a$ and hence $a=q_1\cap\cdots\cap q_n\supset\mathfrak{p}_i>a$, which is impossible. Thus $q_i\not\supset\mathfrak{u}_a$. Let $\mathfrak{c}_i$ be a characteristic ideal associated with q_i. There is always at least one

[9] See §7 for an example.

such ideal, namely, the union of the point ideals of $\mathfrak{L}_a$ divisible by q_i. Now $a=q_1\cap\cdots\cap q_n\supset\mathfrak{c}_1\cap\cdots\cap\mathfrak{c}_n\supset a$ implies $a=\mathfrak{c}_1\cap\cdots\cap\mathfrak{c}_n$. Suppose $\mathfrak{c}_i\supset\mathfrak{c}_1\cap\cdots\cap\mathfrak{c}_{i-1}\cap\mathfrak{c}_{i+1}\cap\cdots\cap\mathfrak{c}_n$. Then $q_i\cap\cdots\cap q_{i-1}\cap q_{i+1}\cap\cdots\cap q_n\supset\mathfrak{p}_i>a$ implies $\mathfrak{c}_1\cap\cdots\cap\mathfrak{c}_{i-1}\cap\mathfrak{c}_{i+1}\cap\cdots\cap\mathfrak{c}_n\supset\mathfrak{p}_i$. But then $a=\mathfrak{c}_i\cap\mathfrak{c}_1\cap\cdots\cap\mathfrak{c}_{i-1}\cap\mathfrak{c}_{i+1}\cap\cdots\cap\mathfrak{c}_n\supset\mathfrak{p}_i$ which is impossible. Hence the representation $a=\mathfrak{c}_1\cap\cdots\cap\mathfrak{c}_n$ is reduced.

Now let $a=\mathfrak{c}_1\cap\cdots\cap\mathfrak{c}_n$ where $\mathfrak{c}_1,\cdots,\mathfrak{c}_n$ are characteristic ideals and the representation is reduced. Let $q_1,\cdots,q_n$ be associated irreducibles. Suppose $q_1\cap\cdots\cap q_n\supset\mathfrak{p}>a$. Then $a=\mathfrak{c}_1\cap\cdots\cap\mathfrak{c}_n\supset\mathfrak{p}>a$ which is impossible. Hence $a=q_1\cap\cdots\cap q_n$. It follows easily that this representation is reduced.

The characteristic ideals of $\mathfrak{L}_a$ can be characterized in terms of the structure of $\mathfrak{L}$ as follows:

THEOREM 3.4. *Let $\mathfrak{S}$ be a Birkhoff lattice in which each element can be expressed as a cross-cut of irreducibles. Then if $\mathfrak{L}_a$ is archimedean, $\mathfrak{c}$ is characteristic if and only if there exists an ideal $\mathfrak{x}\in\mathfrak{L}$ such that $\mathfrak{x}\supset\mathfrak{c}$, $\mathfrak{c}'\cup\mathfrak{x}>\mathfrak{x}$ and $\mathfrak{c}'\cap\mathfrak{x}=a$ for every $\mathfrak{c}'$.*

Proof. Let us first assume that such an ideal $\mathfrak{x}$ exists. Then $\mathfrak{u}_a\cup\mathfrak{x}=\mathfrak{c}\cup\mathfrak{c}'\cup\mathfrak{x}=\mathfrak{c}'\cup\mathfrak{x}$. Let q be an irreducible such that $q\supset\mathfrak{x}$, $q\not\supset\mathfrak{u}_a\cup\mathfrak{x}$ (Lemma 2.6). Since $q\supset\mathfrak{x}\supset\mathfrak{c}$, q divides every point ideal of $\mathfrak{L}_a$ which $\mathfrak{c}$ divides. Now let $q\supset\mathfrak{p}$. Then if $\mathfrak{x}\not\supset\mathfrak{p}$ we have $\mathfrak{c}'\cup\mathfrak{x}=\mathfrak{u}_a\cup\mathfrak{x}\supset\mathfrak{p}\cup\mathfrak{x}\supset\mathfrak{x}$ and $\mathfrak{p}\cup\mathfrak{x}\neq\mathfrak{x}$. Hence $\mathfrak{c}'\cup\mathfrak{x}=\mathfrak{p}\cup\mathfrak{x}$ and $q\supset\mathfrak{p}\cup\mathfrak{x}\supset\mathfrak{c}'\cup\mathfrak{x}$ which contradicts the definition of q. Hence $\mathfrak{x}\supset\mathfrak{p}$. Now if $\mathfrak{c}\not\supset\mathfrak{p}$, then $\mathfrak{c}'\supset\mathfrak{p}$ for some $\mathfrak{c}'$. But then $a=\mathfrak{c}'\cap\mathfrak{x}\supset\mathfrak{p}$ which is impossible. Hence $q\supset\mathfrak{p}$ implies $\mathfrak{c}\supset\mathfrak{p}$ and $\mathfrak{c}$ is thus characteristic.

On the other hand let $\mathfrak{c}$ be characteristic and let q be an irreducible associated with $\mathfrak{c}$. Then $q\cup\mathfrak{c}'>q$ for every $\mathfrak{c}'$. For there is a point ideal $\mathfrak{p}$ such that $\mathfrak{c}'\supset\mathfrak{p}$, $\mathfrak{c}\not\supset\mathfrak{p}$ since otherwise we would have $\mathfrak{c}'=a$ and $\mathfrak{c}=\mathfrak{u}_a$ contrary to the definition of a characteristic ideal. Now $q\cup\mathfrak{p}=q\cup\mathfrak{u}_a>q$ since q is irreducible in $\mathfrak{L}$ by Lemma 2.5. Hence $q\cup\mathfrak{u}_a=q\cup\mathfrak{c}'=q\cup\mathfrak{p}>q$. Now if $\mathfrak{c}'\cap q\neq a$, then $\mathfrak{c}'\cap q\supset\mathfrak{p}>a$ and hence $\mathfrak{c}'\supset\mathfrak{p}$, $q\supset\mathfrak{p}$ by Theorem 2.2. But then $\mathfrak{c}\supset\mathfrak{p}$ and hence $a=\mathfrak{c}\cap\mathfrak{c}'\supset\mathfrak{p}$ which is impossible. Thus $\mathfrak{c}'\cap q=a$ for every $\mathfrak{c}'$.

COROLLARY 3.1. *Each simple ideal of $\mathfrak{L}_a$ is characteristic.*

We may take $\mathfrak{x}$ to be the simple ideal itself.

THEOREM 3.5. *Let $\mathfrak{S}$ be a Birkhoff lattice in which each element can be expressed as a cross-cut of irreducibles. Then if $\mathfrak{L}_a$ is archimedean, each characteristic ideal $\mathfrak{c}$ of $\mathfrak{L}_a$ occurs in a reduced representation $a=\mathfrak{c}\cap\mathfrak{c}_1\cap\cdots\cap\mathfrak{c}_k$ where k is the number of maximal independent point ideals divisible by $\mathfrak{c}$ and $\mathfrak{c}_1,\cdots,\mathfrak{c}_k$ are characteristic ideals of $\mathfrak{L}_a$.*

Proof. Let $\mathfrak{p}_1,\cdots,\mathfrak{p}_k$ be a maximal independent set of point ideals of $\mathfrak{L}_a$ divisible by $\mathfrak{c}$. Imbed $\mathfrak{p}_1,\cdots,\mathfrak{p}_k$ in a maximal independent set $\mathfrak{p}_1,\cdots,\mathfrak{p}_k,\cdots,\mathfrak{p}_n$. Let $\mathfrak{c}_i=\mathfrak{p}_1\cup\cdots\cup\mathfrak{p}_{i-1}\cup\mathfrak{p}_{i+1}\cup\cdots\cup\mathfrak{p}_k\cup\cdots\cup\mathfrak{p}_n$, $i=1,\cdots,k$. If $\mathfrak{c}\cap\mathfrak{c}_1$

$\cap \cdots \cap \mathfrak{c}_k \neq a$ we have $\mathfrak{c} \cap \mathfrak{c}_1 \cap \cdots \cap \mathfrak{c}_k \supset \mathfrak{p} > a$ and $\mathfrak{c} \supset \mathfrak{p}$ implies $\mathfrak{p}_1 \cup \cdots \cup \mathfrak{p}_k \supset \mathfrak{p}$. But then $a = (\mathfrak{p}_1 \cup \cdots \cup \mathfrak{p}_k) \cap \mathfrak{c}_1 \cap \cdots \cap \mathfrak{c}_k \supset \mathfrak{p}$ which is impossible. Hence $a = \mathfrak{c} \cap \mathfrak{c}_1 \cap \cdots \cap \mathfrak{c}_k$. Also since $\mathfrak{c} \cap \mathfrak{c}_1 \cap \cdots \cap \mathfrak{c}_{i-1} \cap \mathfrak{c}_{i+1} \cap \cdots \cap \mathfrak{c}_k \supset \mathfrak{p}_i$ the representation is reduced. Since $\mathfrak{c}_1, \cdots, \mathfrak{c}_k$ are simple ideals of $\mathfrak{L}_a$, they are characteristic by Corollary 3.1.

COROLLARY 3.2. *Let $\mathfrak{S}$ be a Birkhoff lattice in which every element can be expressed as a cross-cut of irreducibles. Then if $\mathfrak{L}_a$ is archimedean of length k, a has a reduced decomposition into irreducibles with k components.*

For by Lemma 1.2 and Theorem 3.4, a has a reduced representation as a cross-cut of k characteristic ideals of $\mathfrak{L}_a$.

LEMMA 3.2. *Let $\mathfrak{S}$ be a Birkhoff lattice and let $\mathfrak{L}_a$ be archimedean for some a. Then $\mathfrak{L}_a$ is modular if and only if it satisfies* B2.

For let $\mathfrak{L}_a$ satisfy B2 and let $\mathfrak{q}$ be a union irreducible ideal of $\mathfrak{L}_a$. If $\mathfrak{s} \not\supset \mathfrak{q}$ and $\mathfrak{s}$ is a simple ideal of $\mathfrak{L}_a$ we have $\mathfrak{q} > \mathfrak{q} \cap \mathfrak{s}$ by B2. Hence since $\mathfrak{q}$ is union irreducible we have $\mathfrak{q} \cap \mathfrak{s} = \mathfrak{q} \cap \mathfrak{s}'$ for any two simple ideals $\mathfrak{s}$ and $\mathfrak{s}'$ which do not divide $\mathfrak{q}$. Let $a = \mathfrak{s}_1 \cap \cdots \cap \mathfrak{s}_n$ where $\mathfrak{s}_1, \cdots, \mathfrak{s}_l \supset \mathfrak{q}$; $\mathfrak{s}_{l+1}, \cdots, \mathfrak{s}_n \not\supset \mathfrak{q}$. Then $a = \mathfrak{q} \cap a = \mathfrak{q} \cap \mathfrak{s}_1 \cap \cdots \cap \mathfrak{s}_n = (\mathfrak{q} \cap \mathfrak{s}_{l+1}) \cap \cdots \cap (\mathfrak{q} \cap \mathfrak{s}_n) = \mathfrak{q} \cap \mathfrak{s}_{l+1} < \mathfrak{q}$. Hence $\mathfrak{q}$ is a point of $\mathfrak{L}_a$ and every ideal of $\mathfrak{L}_a$ is a union of point ideals. Now let $\mathfrak{a} > \mathfrak{a} \cap \mathfrak{b}$ in $\mathfrak{L}_a$. Then since every ideal is a union of point ideals, there exists a point ideal $\mathfrak{p}$ such that $\mathfrak{a} \supset \mathfrak{p}$, $\mathfrak{a} \cap \mathfrak{b} \not\supset \mathfrak{p}$. But then $\mathfrak{a} = (\mathfrak{a} \cap \mathfrak{b}) \cup \mathfrak{p}$. Hence $\mathfrak{a} \cup \mathfrak{b} = (\mathfrak{a} \cap \mathfrak{b}) \cup \mathfrak{p} \cup \mathfrak{b} = \mathfrak{p} \cup \mathfrak{b} > \mathfrak{b}$ since $\mathfrak{S}$ is a Birkhoff lattice. Thus B1 and B2 hold in $\mathfrak{L}_a$ and $\mathfrak{L}_a$ is modular by Lemma 1.5. Conversely, if $\mathfrak{L}_a$ is modular, then B2 is satisfied by Lemma 1.5. This completes the proof.

According to Theorem 3.1, if every element of a lattice $\mathfrak{S}$ has a decomposition into irreducibles and the number of components in the decompositions of a is bounded, then $\mathfrak{L}_a$ is archimedean. This result can be sharpened considerably if $\mathfrak{S}$ is modular.

LEMMA 3.3. *Let $\mathfrak{S}$ be a modular lattice. Then if an element a has a decomposition into irreducibles, $\mathfrak{L}_a$ is archimedean.*

For let $a = q_1 \cap \cdots \cap q_k$ where $q_1, \cdots, q_k$ are irreducible. Since $\mathfrak{S}$ is modular, $\mathfrak{L}$ is modular by Lemma 2.2. Now if $q_i \not\supset \mathfrak{p}$ where $\mathfrak{p} > a$, we have $q_i \cup \mathfrak{p} > q_i$ and hence $q_i \cup \mathfrak{u}_a > q_i$ since q_i is irreducible. But then $\mathfrak{u}_a > \mathfrak{u}_a \cap q_i$ since $\mathfrak{L}$ is modular. Thus each irreducible q_i divides a simple characteristic ideal $\mathfrak{c}_i = q_i \cap \mathfrak{u}_a$. Since $\mathfrak{L}$ is modular, we have $\mathfrak{u}_a > \mathfrak{c}_1 > \mathfrak{c}_1 \cap \mathfrak{c}_2 > \cdots > \mathfrak{c}_1 \cap \cdots \cap \mathfrak{c}_k = a$. Hence $\mathfrak{L}_a$ is archimedean and the lemma is proved.

If $\mathfrak{S}$ is modular and a has two reduced decompositions into irreducibles, then by Lemma 3.3, $\mathfrak{L}_a$ is archimedean and a has two reduced representations as a cross-cut of simple ideals. Now by Lemma 3.2, B2 holds in $\mathfrak{L}_a$ and hence by the dual of Lemma 1.3 any two reduced representations of a as a cross-cut of simple ideals have the same number of components and any simple ideal

of one decomposition may be replaced by a suitably chosen simple ideal of the other. Thus by Theorem 3.3 and Corollary 3.1 we have the

KUROSCH-ORE DECOMPOSITION THEOREM. *Let an element of a modular lattice have two reduced decompositions into irreducibles. Then the number of components in the two decompositions is the same and any component in one decomposition may be replaced by a suitably chosen component of the other.*

4. Lattices with unique decompositions. This section will be devoted to the proof of the following theorem:

THEOREM 4.1. *Let $\mathfrak{S}$ satisfy the ascending chain condition. Then each element of $\mathfrak{S}$ has a unique representation as a reduced cross-cut of irreducibles if and only if $\mathfrak{S}$ is a Birkhoff lattice and $\mathfrak{L}_a$ is a Boolean algebra for each a.*

We begin with a series of lemmas, the first of which proves the necessity of the conditions of the theorem.

LEMMA 4.1. *Let $\mathfrak{S}$ satisfy the ascending chain condition and let each element have a unique representation as a reduced cross-cut of irreducibles. Then $\mathfrak{S}$ is a Birkhoff lattice and $\mathfrak{L}_a$ is a Boolean algebra for each a.*

For let $\mathfrak{b}>a$, $\mathfrak{c}\supset a$ and $\mathfrak{c}\not\supset\mathfrak{b}$. If $\mathfrak{b}\cup\mathfrak{c}\not>\mathfrak{c}$ we have $\mathfrak{b}\cup\mathfrak{c}\supset\mathfrak{d}\supset\mathfrak{c}$ where $\mathfrak{b}\cup\mathfrak{c}\neq\mathfrak{d}\neq\mathfrak{c}$. Since $\mathfrak{d}\not\supset\mathfrak{b}$, there exists a $d\in\mathfrak{d}$ such that $d\not\supset\mathfrak{b}$. Since $\mathfrak{d}\neq\mathfrak{c}$, there exists a c such that $c\in\mathfrak{c}$, $d\supset c$, and $c\not\supset\mathfrak{d}$. Furthermore since $c\not\supset\mathfrak{d}$ there exists an irreducible q_c such that $q_c\supset c$, $q_c\not\supset\mathfrak{d}$ (Lemma 2.6). But then $\mathfrak{b}\supset\mathfrak{b}\cap q_c\supset a$ and if $\mathfrak{b}=\mathfrak{b}\cap q_c$ we have $q_c\supset\mathfrak{b}\cup\mathfrak{c}\supset\mathfrak{d}$ which contradicts $q_c\not\supset\mathfrak{d}$. Hence $a=\mathfrak{b}\cap q_c$. Similarly there exists an irreducible q_d such that $q_d\supset d$ and $a=\mathfrak{b}\cap q_d$. By Theorem 2.2 we have $a=b_c\cap q_c$ and $a=b_d\cap q_d$ where b_c, $b_d\in\mathfrak{b}$. Let $b=b_c\cap b_d$. Then $b\in\mathfrak{b}$ and $a=b\cap q_c=b\cap q_d$. Let $b=q_1\cap\cdots\cap q_k$. Then a has two reduced representations $a=q_{i_1}\cap\cdots\cap q_{i_l}\cap q_c=q_{j_1}\cap\cdots\cap q_{j_m}\cap q_d$. Now $q_c\neq q_d$ since otherwise $q_c\supset\mathfrak{d}$ and $q_c\neq q_{j_r}$ since otherwise $q_c\supset\mathfrak{b}\cup\mathfrak{c}\supset\mathfrak{d}$ contrary to $q_c\not\supset\mathfrak{d}$. Hence a has two distinct reduced representations as a cross-cut of irreducibles which contradicts our hypothesis. Thus $\mathfrak{b}\cup\mathfrak{c}>\mathfrak{c}$ and hence each element of $\mathfrak{S}$ satisfies the Birkhoff condition in the lattice of ideals.

Now since each element has a unique decomposition into irreducibles, the number of components is obviously bounded and hence $\mathfrak{L}_a$ is archimedean by Theorem 3.1. Let $\mathfrak{p}_1,\cdots,\mathfrak{p}_k$ be a maximal independent set of point ideals of $\mathfrak{L}_a$. Then $\mathfrak{p}_1,\cdots,\mathfrak{p}_k$ generate a Boolean algebra with simple ideals $\mathfrak{s}_1,\cdots,\mathfrak{s}_k$. $\mathfrak{s}_1,\cdots,\mathfrak{s}_k$ are clearly simple ideals of $\mathfrak{L}_a$ and hence are characteristic ideals by Corollary 3.2. Thus a has a decomposition $a=q_1\cap\cdots\cap q_k$ where $q_i\supset\mathfrak{s}_i$ (Theorem 3.3). Now suppose there is a simple ideal $\mathfrak{s}$ distinct from $\mathfrak{s}_1,\cdots,\mathfrak{s}_k$. Let $q\supset\mathfrak{s}$, $q\not\supset\mathfrak{u}_a$. Then q is a component of a by Theorem 3.5 and hence $q=q_i$ for some i since a has but one reduced decomposition into irreducibles. But then $q\supset\mathfrak{s}\cup\mathfrak{s}_i=\mathfrak{u}_a$ which is impossible. Hence $\mathfrak{s}_1,\cdots,\mathfrak{s}_k$ are all of the simple

ideals of $\mathfrak{L}_a$ and since each ideal of $\mathfrak{L}_a$ can be expressed as a union cross-cut of simple ideals, $\mathfrak{L}_a$ is simply the Boolean algebra generated by $\mathfrak{p}_1, \cdots, \mathfrak{p}_k$.

LEMMA 4.2. *If $\mathfrak{L}_a$ is a Boolean algebra, then it is archimedean.*

For if $\mathfrak{L}_a$ has an infinite number of point ideals, let $\mathfrak{p}_1, \mathfrak{p}_2, \mathfrak{p}_3, \cdots$ be a denumerable sequence of point ideals. Let $\mathfrak{p}_i' = \mathfrak{p}_1 \cup \mathfrak{p}_2 \cup \cdots \cup \mathfrak{p}_{i-1} \cup \mathfrak{p}_{i+1} \cup \cdots$. Then since $\mathfrak{L}_a$ is a Boolean algebra we have $a = \mathfrak{p}_1' \cap \mathfrak{p}_2' \cap \cdots$. But since the cross-cut of an infinite number of ideals consists of all elements contained in finite cross-cuts $a = \mathfrak{p}_1' \cap \mathfrak{p}_2' \cap \cdots \cap \mathfrak{p}_k'$ for some k. Then $a \supset \mathfrak{p}_{k+1}$ which contradicts $\mathfrak{p}_{k+1} > a$. Hence $\mathfrak{L}_a$ has only a finite number of point ideals and thus is archimedean.

LEMMA 4.3. *Let $\mathfrak{S}$ be a Birkhoff lattice in which each $\mathfrak{L}_a$ is archimedean. Then if every three ideals covering a principal ideal generate a Boolean algebra of order eight, $\mathfrak{L}_a$ is a Boolean algebra for each a.*

For let the hypotheses of the lemma be satisfied and let every three ideals covering a principal ideal generate a Boolean algebra. We show first that the ideals of any finite set of ideals covering a principal ideal are independent. Suppose that for any a every $k-1$ ideals covering a are independent. Let $\mathfrak{p}_1, \cdots, \mathfrak{p}_k$ be k distinct ideals covering a. If $\mathfrak{p}_1, \cdots, \mathfrak{p}_k$ are not independent let $\mathfrak{p}_1 \cup \mathfrak{p}_2 \cup \cdots \cup \mathfrak{p}_{k-1} \supset \mathfrak{p}_k$ say. Now $\mathfrak{p}_1 \cup \mathfrak{p}_i \not\supset \mathfrak{p}_2, \cdots, \mathfrak{p}_{i-1}, \mathfrak{p}_{i+1}, \cdots, \mathfrak{p}_k$ $(i=2, \cdots, k)$ since every three ideals covering a generate a Boolean algebra. Hence elements $x_{ij} \in \mathfrak{p}_1$ exist such that $x_{ij} \cup \mathfrak{p}_i \not\supset \mathfrak{p}_j$ $(j=2, \cdots, i-1, i+1, \cdots, k;$ $i=2, \cdots, k)$. Let $x = x_{23} \cap x_{24} \cap \cdots \cap x_{k\,k-1}$. Then $x \in \mathfrak{p}_1$ and $x \cup \mathfrak{p}_i \not\supset \mathfrak{p}_2, \cdots, \mathfrak{p}_{i-1}, \mathfrak{p}_{i+1}, \cdots, \mathfrak{p}_k$ $(i=2, \cdots, k)$. Clearly $x \not\supset \mathfrak{p}_2, \cdots, \mathfrak{p}_k$. Hence $\mathfrak{p}_2' = x \cup \mathfrak{p}_2 > x, \cdots, \mathfrak{p}_k' = x \cup \mathfrak{p}_k > x$ and $\mathfrak{p}_2', \cdots, \mathfrak{p}_k'$ are distinct. Thus by the induction assumption $\mathfrak{p}_2', \cdots, \mathfrak{p}_k'$ are independent. But $\mathfrak{p}_2' \cup \cdots \cup \mathfrak{p}_{k-1}' \supset x \cup \mathfrak{p}_2 \cup \cdots \cup \mathfrak{p}_{k-1} \supset x \cup \mathfrak{p}_1 \cup \cdots \cup \mathfrak{p}_{k-1} \supset x \cup \mathfrak{p}_k = \mathfrak{p}_k'$ which is contrary to the independence. Hence the independence of any finite set of covering elements follows by induction.

Now let $\mathfrak{a} \in \mathfrak{L}_a$ and let $\mathfrak{p}_1, \cdots, \mathfrak{p}_k$ be a maximal independent set of point ideals of $\mathfrak{L}_a$ divisible by $\mathfrak{a}$. Imbed $\mathfrak{p}_1, \cdots, \mathfrak{p}_k$ in a maximal independent set $\mathfrak{p}_1, \cdots, \mathfrak{p}_k, \cdots, \mathfrak{p}_n$. Set $\mathfrak{b} = \mathfrak{p}_1 \cup \cdots \cup \mathfrak{p}_k$. Then $\mathfrak{a} \supset \mathfrak{b}$. If $\mathfrak{b} \not\supset \mathfrak{a}$, there exists an element $b_1 \in \mathfrak{b}$ such that $b_1 \not\supset \mathfrak{a}$. Now $\mathfrak{b} \cup \mathfrak{p}_{k+i} \not\supset \mathfrak{p}_{k+j}$ $(j=k+1, \cdots, k+i-1,$ $k+i+1, \cdots, n;\ i=1, \cdots, n-k)$. Hence as above there exists an element $b_2 \in \mathfrak{b}$ such that $b_2 \cup \mathfrak{p}_{k+i} \not\supset \mathfrak{p}_{k+j}$, $i \neq j$. Also $\mathfrak{a} = \mathfrak{a} \cup \mathfrak{b} \not\supset \mathfrak{p}_{k+1}, \cdots, \mathfrak{p}_n$. Hence an element $b_3 \in \mathfrak{b}$ exists such that $\mathfrak{a} \cup b_3 \not\supset \mathfrak{p}_{k+1}, \cdots, \mathfrak{p}_n$. Set $b = b_1 \cap b_2 \cap b_3$. Then $b \in \mathfrak{b}$, $b \not\supset \mathfrak{a}$, $b \cup \mathfrak{p}_{k+i} \not\supset \mathfrak{p}_{k+j}$, $i \neq j$, and $\mathfrak{a} \cup b \not\supset \mathfrak{p}_{k+1}, \cdots, \mathfrak{p}_n$. Clearly $b \not\supset \mathfrak{p}_{k+1}, \cdots, \mathfrak{p}_n$. Hence $\mathfrak{p}_{k+1}' = b \cup \mathfrak{p}_{k+1} > b, \cdots, \mathfrak{p}_n' = b \cup \mathfrak{p}_n > b$ and $\mathfrak{p}_{k+1}', \cdots, \mathfrak{p}_n'$ are distinct. Let $b \cup \mathfrak{a} \supset \mathfrak{p} > b$. Then $\mathfrak{p}$ is distinct from $\mathfrak{p}_{k+1}', \cdots, \mathfrak{p}_n'$. For if $\mathfrak{p} = \mathfrak{p}_{k+i}'$, then $b \cup \mathfrak{a} \supset \mathfrak{p}_{k+i}$ contrary to the definition of b. Thus by the result of the above paragraph $\mathfrak{p}, \mathfrak{p}_{k+1}', \cdots, \mathfrak{p}_n'$ are independent. But $\mathfrak{p}_{k+1}' \cup \cdots \cup \mathfrak{p}_n' = b \cup \mathfrak{p}_{k+1} \cup \cdots \cup \mathfrak{p}_n = b \cup \mathfrak{p}_1 \cup \cdots \cup \mathfrak{p}_n \supset b \cup \mathfrak{a} \supset \mathfrak{p}$ which is impossible. Hence $\mathfrak{a} = \mathfrak{b}$ and

$\mathfrak{L}_a$ is a point lettice. But then the point ideals of $\mathfrak{L}_a$ are independent and generate $\mathfrak{L}_a$. Thus $\mathfrak{L}_a$ is a Boolean algebra by Lemma 1.2.

LEMMA 4.4. *Let $\mathfrak{S}$ be a Birkhoff lattice satisfying the ascending chain condition in which every three ideals covering a principal ideal generate a Boolean algebra. Let q be an irreducible of $\mathfrak{S}$ such that $q \supset a$; $\mathfrak{b}$, $\mathfrak{c} > a$ and $\mathfrak{b} \neq \mathfrak{c}$. Then either $q \supset \mathfrak{b}$ or $q \supset \mathfrak{c}$.*

Let us suppose that for some a we have $q \supset a$; $\mathfrak{b}$, $\mathfrak{c} > a$, $\mathfrak{b} \neq \mathfrak{c}$, $q \not\supset \mathfrak{b}$ and $q \not\supset \mathfrak{c}$. We shall show that a proper divisor a' of a exists with the same properties and hence the lemma follows from the ascending chain condition. Now $q \neq a$ since otherwise $q = \mathfrak{b} \cap \mathfrak{c}$ contrary to the irreducibility of q. Hence $q \supset \mathfrak{p} > a$ by Theorem 2.1. Clearly $\mathfrak{p} \neq \mathfrak{b}$, $\mathfrak{c}$ since otherwise $q \supset \mathfrak{b}$ or $q \supset \mathfrak{c}$. Hence $\mathfrak{p}$, $\mathfrak{b}$ and $\mathfrak{c}$ generate a Boolean algebra. Since $\mathfrak{p} \cup \mathfrak{b} \not\supset \mathfrak{c}$ there exists an element $p \in \mathfrak{p}$ such that $p \cup \mathfrak{b} \not\supset \mathfrak{c}$. Since $q \supset \mathfrak{p}$, there exists an element $p' \in \mathfrak{p}$ such that $q \supset p'$. Let $a' = p \cap p'$. Then $a' \in \mathfrak{p}$ and hence $a' \neq a$. Clearly $q \supset a'$. Let $\mathfrak{b}' = a' \cup \mathfrak{b}$, $\mathfrak{c}' = a' \cup \mathfrak{c}$. Then $\mathfrak{b}' > a'$ and $\mathfrak{c}' > a'$ by the Birkhoff condition. If $\mathfrak{b}' = \mathfrak{c}'$, then $p \cup \mathfrak{b} \supset a' \cup \mathfrak{b} \supset \mathfrak{c}$, which contradicts $p \cup \mathfrak{b} \not\supset \mathfrak{c}$. Hence $\mathfrak{b}' \neq \mathfrak{c}'$. Since $q \not\supset \mathfrak{b}$, $q \not\supset \mathfrak{c}$ we have $q \not\supset \mathfrak{b}'$, $q \not\supset \mathfrak{c}'$. Thus a' is a proper divisor of a with the desired properties.

LEMMA 4.5. *Let $\mathfrak{S}$ be a Birkhoff lattice satisfying the ascending chain condition in which every three ideals covering a principal ideal generate a Boolean algebra. Then if a has a reduced representation $a = q_1 \cap \cdots \cap q_k$, $\mathfrak{L}_a$ is archimedean of length k and each q_i divides a simple ideal of $\mathfrak{L}_a$.*

For let $\mathfrak{a}_i$ be the union of the point ideals of $\mathfrak{L}_a$ which are divisible by q_i. Then $\mathfrak{a}_i \neq \mathfrak{u}_a$ since $\mathfrak{a}_i$ is a characteristic ideal of $\mathfrak{L}_a$. Now let $\mathfrak{p}$, $\mathfrak{p}'$ be any two point ideals of $\mathfrak{L}_a$ which are not divisible by $\mathfrak{a}_i$. Then $\mathfrak{a}_i \cup \mathfrak{p} > \mathfrak{a}_i$ and $\mathfrak{a}_i \cup \mathfrak{p}' > \mathfrak{a}_i$ by the Birkhoff condition. Now suppose that $\mathfrak{a}_i \cup \mathfrak{p} \not\supset \mathfrak{p}'$. Then there exists an element $a_1 \in \mathfrak{a}_i$ such that $a_1 \cup \mathfrak{p} \not\supset \mathfrak{p}'$. Since $q_i \supset \mathfrak{a}_i$, there exists an element $a_2 \in \mathfrak{a}_i$ such that $q_i \supset a_2$. Let $a_i = a_1 \cap a_2$. Then $q_i \supset a_i$ and $a_i \cup \mathfrak{p} \not\supset \mathfrak{p}'$. Clearly $a_i \not\supset \mathfrak{p}'$. If $a_i \supset \mathfrak{p}$, then $q_i \supset \mathfrak{p}$ and $\mathfrak{a}_i \supset \mathfrak{p}$ contrary to assumption. Hence $a_i \cup \mathfrak{p} > a_i$, $a_i \cup \mathfrak{p}' > a_i$ and $a_i \cup \mathfrak{p} \neq a_i \cup \mathfrak{p}'$. Since $q_i \supset a_i$ by Lemma 4.4 we have either $q_i \supset a_i \cup \mathfrak{p}$ or $q_i \supset a_i \cup \mathfrak{p}'$. Hence $q_i \supset \mathfrak{p}$ or $q_i \supset \mathfrak{p}'$. But then $\mathfrak{a}_i \supset \mathfrak{p}$ or $\mathfrak{a}_i \supset \mathfrak{p}'$ contrary to assumption. Thus $\mathfrak{a}_i \cup \mathfrak{p} \supset \mathfrak{p}'$ and $\mathfrak{a}_i \cup \mathfrak{p} = \mathfrak{a}_i \cup \mathfrak{p}'$ for every pair of point ideals of $\mathfrak{L}_a$ not divisible by $\mathfrak{a}_i$. But then $\mathfrak{a}_i \cup \mathfrak{p} = \mathfrak{a}_i \cup \mathfrak{u}_a = \mathfrak{u}_a$ and $\mathfrak{u}_a > \mathfrak{a}_i$. Hence $\mathfrak{a}_i$ is simple and each q_i divides a simple ideal of $\mathfrak{L}_a$.

Now let $\mathfrak{b}_0 = \mathfrak{u}_a$ and let $\mathfrak{b}_i$ denote the union of the point ideals of $\mathfrak{L}_a$ which are divisible by $q_1, \cdots, q_i$. Then $\mathfrak{b}_1 = \mathfrak{a}_1$ and $\mathfrak{b}_0 > \mathfrak{a}_1$ by the result we have just obtained. Clearly $\mathfrak{b}_{l-1} \supset \mathfrak{b}_l$. If $\mathfrak{b}_{l-1} = \mathfrak{b}_l$, let $q_1 \cap \cdots \cap q_{l-1} \cap q_{l+1} \cap \cdots \cap q_k \supset \mathfrak{p}_l > a$. $\mathfrak{p}_l$ exists since the representation is reduced. Now $q_1 \cap \cdots \cap q_{l-1} \supset \mathfrak{p}_l$ and hence $\mathfrak{b}_{l-1} \supset \mathfrak{p}_l$. But then $\mathfrak{b}_l \supset \mathfrak{p}_l$ and hence $q_l \supset \mathfrak{p}_l$. Thus $a = q_1 \cap \cdots \cap q_k \supset \mathfrak{p}_l$ which is impossible. Hence $\mathfrak{b}_{l-1} \neq \mathfrak{b}_l$. Now let $\mathfrak{p}$ and $\mathfrak{p}'$ be two point ideals divisible by $\mathfrak{b}_{l-1}$ but not by $\mathfrak{b}_l$. If $\mathfrak{b}_l \cup \mathfrak{p} \neq \mathfrak{b}_l \cup \mathfrak{p}'$ there exists an element $b_l \in \mathfrak{b}_l$ such that $q_l \supset b_l$, $b_l \cup \mathfrak{p} > b_l$, $b_l \cup \mathfrak{p}' > b_l$ and $b_l \cup \mathfrak{p} \neq b_l \cup \mathfrak{p}'$. But then $q_l \supset b_l \cup \mathfrak{p}$

or $q_l \supset b_l \cup \mathfrak{p}'$ by Lemma 4.4. Hence either $\mathfrak{b}_l \supset \mathfrak{p}$ or $\mathfrak{b}_l \supset \mathfrak{p}'$ which is contrary to assumption. Hence $\mathfrak{b}_l \cup \mathfrak{p} = \mathfrak{b}_l \cup \mathfrak{p}'$ for every two point ideals of $\mathfrak{b}_{l-1}$ which are not divisible by $\mathfrak{b}_l$. Thus $\mathfrak{b}_l \cup \mathfrak{p} = \mathfrak{b}_l \cup \mathfrak{b}_{l-1} = \mathfrak{b}_{l-1}$ and $\mathfrak{b}_{l-1} > \mathfrak{b}_l$ by the Birkhoff condition. Hence we have the chain $\mathfrak{u}_a > \mathfrak{b}_1 > \mathfrak{b}_2 > \cdots > \mathfrak{b}_k$. But $\mathfrak{b}_k = a$ and the lemma follows from Lemma 3.1.

LEMMA 4.6. *Let $\mathfrak{S}$ be a Birkhoff lattice satisfying the ascending chain condition in which every three ideals covering a principal ideal generate a Boolean algebra. Then each $a \in \mathfrak{S}$ has a unique reduced representation $a = q_1 \cap \cdots \cap q_k$ where $q_1, \cdots, q_k$ are irreducibles. $\mathfrak{L}_a$ is a Boolean algebra of order 2^k and each q_i divides a simple ideal of $\mathfrak{L}_a$.*

It follows from Lemmas 4.3 and 4.5 that $\mathfrak{L}_a$ is a Boolean algebra of order 2^k. q_i divides a simple ideal $\mathfrak{s}_i$ of $\mathfrak{L}_a$ by Lemma 4.5. Now let $a = q_1' \cap \cdots \cap q_l'$ be a reduced decomposition of a. By Lemma 4.5, $l = k$ and q_i' divides a simple ideal $\mathfrak{s}_j$. Let $b = q_i' \cap q_j$. Then $b \supset \mathfrak{s}_j$, and $b \not\supset \mathfrak{u}_a$. Let $\mathfrak{s}_j \not\supset \mathfrak{p} > a$. Then $u_a = \mathfrak{s}_j \cup \mathfrak{p}$ and $b \cup \mathfrak{p} = b \cup \mathfrak{s}_j \cup \mathfrak{p} = b \cup \mathfrak{u}_a > b$ by the Birkhoff condition. If $q_j \neq b$, we have $q_j \supset \mathfrak{p}_j > b$ and $\mathfrak{p}_j \neq \mathfrak{p} \cup b$ since otherwise $q_j \supset \mathfrak{p} \cup \mathfrak{s}_j = \mathfrak{u}_a$. Hence by Lemma 4.4, either $q_i' \supset \mathfrak{p}_j$ or $q_i' \supset \mathfrak{p}$. But if $q_i' \supset \mathfrak{p}_j$, then $b = q_i' \cap q_j \supset \mathfrak{p}_j > b$ which is impossible. Hence $q_i' \supset \mathfrak{p}$ and $q_i' \supset \mathfrak{u}_a$ which is impossible. Thus $q_j = b$ and similarly $q_i' = b$. Hence q_i' is equal to q_j and the two representations are identical. This completes the proof of the lemma.

Lemma 4.1 and Lemma 4.6 together give Theorem 4.1.

In view of Lemma 4.6, lattices with unique irreducible decompositions may be characterized in terms of the local properties of the lattice of ideals as follows:

THEOREM 4.2. *Let $\mathfrak{S}$ satisfy the ascending chain condition. Then each element of $\mathfrak{S}$ has a unique reduced decomposition into irreducibles if and only if $\mathfrak{S}$ is a Birkhoff lattice in which every three ideals covering an element of $\mathfrak{S}$ are independent.*

As a corollary to Lemma 4.6 we have

COROLLARY 4.1. *Let $\mathfrak{S}$ satisfy the ascending chain condition and let every element of $\mathfrak{S}$ have a unique reduced decomposition into irreducibles. Then the number of irreducible components of a is equal to the number of ideals covering a.*

COROLLARY 4.2. *Let $\mathfrak{S}$ be a Birkhoff lattice satisfying the ascending chain condition. Then if $\mathfrak{S}$ contains a modular, non-distributive sublattice, the lattice of ideals of $\mathfrak{S}$ contains a complete*(10) *modular, non-distributive sublattice of order five.*

For if $\mathfrak{S}$ contains a modular, non-distributive sublattice of order five, at

---

(10) A sublattice $\mathfrak{L}'$ of $\mathfrak{L}$ is said to be *complete* if $\mathfrak{a} > \mathfrak{b}$ in $\mathfrak{L}'$ implies $\mathfrak{a} > \mathfrak{b}$ in $\mathfrak{L}$.

least one element of $\mathfrak{S}$ does *not* have a unique decomposition into irreducibles. But then there are three ideals covering a principal ideal which are dependent. These three ideals generate a complete, modular, non-distributive sublattice of $\mathfrak{L}$ of order five.

5. **Unicity of the number of components.** In the previous section lattices with unique irreducible decompositions were completely characterized as Birkhoff lattices with certain special properties. Simple examples show that a similar characterization of lattices in which the *number* of components is unique will require lattices that are considerably more general than Birkhoff lattices. Hence we shall restrict ourselves to the characterization of Birkhoff lattices having the number of components unique. We prove the following theorem:

THEOREM 5.1. *Let $\mathfrak{S}$ be a Birkhoff lattice satisfying the ascending chain condition. Then the number of components in the reduced decompositions of each element into irreducibles is unique if and only if $\mathfrak{L}_a$ is modular for each a.*

As in §4, the proof rests on a series of lemmas.

LEMMA 5.1. *Let $\mathfrak{S}$ be a Birkhoff lattice satisfying the ascending condition. Then the number of components in the irreducible decompositions of a is unique if and only if $\mathfrak{L}_a$ is archimedean, modular, and every characteristic ideal of $\mathfrak{L}_a$ is simple.*

Since the ascending chain condition holds each element of $\mathfrak{S}$ has a decomposition into irreducibles. Now if the number of components in the irreducible decompositions of a is unique it is certainly bounded and hence $\mathfrak{L}_a$ is archimedean by Theorem 3.1. Now let $\mathfrak{c}$ be a characteristic ideal of $\mathfrak{L}_a$ and let $\mathfrak{p}_1, \cdots, \mathfrak{p}_k$ be a maximal independent set of point ideals divisible by $\mathfrak{c}$. Imbed $\mathfrak{p}_1, \cdots, \mathfrak{p}_k$ in a maximal independent set $\mathfrak{p}_1, \cdots, \mathfrak{p}_k, \cdots, \mathfrak{p}_n$. By Theorem 3.5 and Theorem 3.3, a has an irreducible decomposition having $k+1$ components. But by Corollary 3.2 a has a decomposition having n components. Hence if the number of components is unique we have $n=k+1$. But then $\mathfrak{p}_1 \cup \cdots \cup \mathfrak{p}_k$ is a simple ideal of $\mathfrak{L}_a$ and $\mathfrak{u}_a \supset \mathfrak{c} \supset \mathfrak{p}_1 \cup \cdots \cup \mathfrak{p}_k$, $\mathfrak{u}_a \neq \mathfrak{c}$. Hence $\mathfrak{c} = \mathfrak{p}_1 \cup \cdots \cup \mathfrak{p}_k$ and $\mathfrak{c}$ is a simple ideal of $\mathfrak{L}_a$.

Now let $\mathfrak{s}$ be an arbitrary simple ideal of $\mathfrak{L}_a$ and let $\mathfrak{a}$ be any ideal of $\mathfrak{L}_a$ such that $\mathfrak{s} \not\supset \mathfrak{a}$. By Theorem 3.2, $\mathfrak{a}$ has a reduced representation $\mathfrak{a} = \mathfrak{s}_1 \cap \cdots \cap \mathfrak{s}_l$ where $\mathfrak{s}_1, \cdots, \mathfrak{s}_l$ are simple ideals of $\mathfrak{L}_a$. If $\mathfrak{a} \cap \mathfrak{s} \neq a$, by Theorem 3.2 there exists a simple ideal $\mathfrak{s}_{l+2}$ such that $\mathfrak{s}_{l+2} \not\supset \mathfrak{a} \cap \mathfrak{s}$. Similarly if $\mathfrak{a} \cap \mathfrak{s} \cap \mathfrak{s}_{l+2} \neq a$, there exists a simple ideal $\mathfrak{s}_{l+3}$ such that $\mathfrak{s}_{l+3} \not\supset \mathfrak{a} \cap \mathfrak{s} \cap \mathfrak{s}_{l+2}$. Thus we eventually have $\mathfrak{a} \cap \mathfrak{s} \cap \mathfrak{s}_{l+2} \cap \cdots \cap \mathfrak{s}_m = a$. Then $a = \mathfrak{s}_1 \cap \cdots \cap \mathfrak{s}_l \cap \mathfrak{s} \cap \mathfrak{s}_{l+2} \cap \cdots \cap \mathfrak{s}_m$ and since each simple ideal is characteristic this decomposition gives a decomposition into irreducibles with the same number of terms. Hence if the number of components in the irreducible decompositions of a is unique we have $m \geqq n$ where n is the length of $\mathfrak{L}_a$. But $\mathfrak{u}_a \supset \mathfrak{s}_1 \supset \mathfrak{s}_1 \cap \mathfrak{s}_2 \supset \cdots \supset \mathfrak{s}_1 \cap \cdots \cap \mathfrak{s}_l \supset \mathfrak{s}_1$

$\cap \cdots \cap \mathfrak{s}_l \cap \mathfrak{s} \supset \mathfrak{s}_1 \cap \cdots \cap \mathfrak{s}_l \cap \mathfrak{s} \cap \mathfrak{s}_{l+2} \supset \cdots \supset \mathfrak{s}_1 \cap \cdots \cap \mathfrak{s}_m = a$ and the ideals of this chain are distinct. Hence $m \leqq n$ by Lemma 3.1. Thus $m = n$ and each ideal of the chain covers the ideal which immediately follows. Hence $\mathfrak{a} > \mathfrak{a} \cap \mathfrak{s}$. Now let $\mathfrak{a}$ and $\mathfrak{b}$ be any two ideals of $\mathfrak{L}_a$ such that $\mathfrak{a} \cup \mathfrak{b} > \mathfrak{b}$. By Theorem 3.2 and ideal $\mathfrak{s}$ exists such that $\mathfrak{s} \supset \mathfrak{b}$, $\mathfrak{s} \not\supset \mathfrak{a} \cup \mathfrak{b}$. But then $\mathfrak{b} = (\mathfrak{a} \cup \mathfrak{b}) \cap \mathfrak{s}$. Hence $\mathfrak{a} > \mathfrak{a} \cap \mathfrak{s} = \mathfrak{a} \cap (\mathfrak{a} \cup \mathfrak{b}) \cap \mathfrak{s} = \mathfrak{a} \cap \mathfrak{b}$. Thus $\mathfrak{a} \cup \mathfrak{b} > \mathfrak{b}$ implies $\mathfrak{a} > \mathfrak{a} \cap \mathfrak{b}$ and B2 holds in $\mathfrak{L}_a$. But then $\mathfrak{L}_a$ is modular by Lemma 3.2.

On the other hand let $\mathfrak{L}_a$ be archimedean, modular, and every characteristic ideal be simple. Let $a = q_1 \cap \cdots \cap q_k$ be a reduced decomposition into irreducibles. By Theorem 3.3, a has a reduced representation $a = \mathfrak{c}_1 \cap \cdots \cap \mathfrak{c}_k$ where $\mathfrak{c}_i$ is a characteristic ideal of $\mathfrak{L}_a$. But then $\mathfrak{c}_i$ is a simple ideal of $\mathfrak{L}_a$ by assumption. Thus $\mathfrak{u}_a > \mathfrak{c}_1 > \mathfrak{c}_1 \cap \mathfrak{c}_2 > \cdots > \mathfrak{c}_1 \cap \cdots \cap \mathfrak{c}_k = a$ since B2 holds in $\mathfrak{L}_a$ by Lemma 1.5. Hence k is simply the length of $\mathfrak{L}_a$ and every reduced decomposition of a into irreducibles has the same number of components. This completes the proof of the lemma.

LEMMA 5.2. *Let $\mathfrak{S}$ be a Birkhoff lattice satisfying the ascending chain condition. Then if $\mathfrak{L}_a$ is modular for each a, every characteristic ideal of $\mathfrak{L}_a$ is simple.*

Let every characteristic ideal of $\mathfrak{L}_b$ be simple for every proper divisor b of a. We shall show that every characteristic ideal of $\mathfrak{L}_a$ is simple and the lemma follows by the ascending chain condition.

If $\mathfrak{c}$ is a characteristic ideal of $\mathfrak{L}_a$ which is not simple, let q be an associated irreducible. If x is any element of $\mathfrak{S}$ divisible by q, let $\mathfrak{q}_x$ denote the union of the point ideals of $\mathfrak{L}_x$ divisible by q. Then since $\mathfrak{c}$ is a characteristic ideal associated with q we have $\mathfrak{c} \supset \mathfrak{q}_a$ and hence $\mathfrak{q}_a$ is not a simple ideal of $\mathfrak{L}_a$. Now suppose that for every two point ideals $\mathfrak{p}$ and $\mathfrak{p}'$ such that $\mathfrak{q}_a \not\supset \mathfrak{p}$, $\mathfrak{p}'$ we have $\mathfrak{q}_a \cup \mathfrak{p} = \mathfrak{q}_a \cup \mathfrak{p}'$. Then $\mathfrak{u}_a = \mathfrak{q}_a \cup \mathfrak{u}_a = \mathfrak{q}_a \cup \mathfrak{p} > \mathfrak{q}_a$ and $\mathfrak{q}_a$ is simple contrary to assumption. Hence there are two point ideals $\mathfrak{p}$ and $\mathfrak{p}'$ such that $\mathfrak{q}_a \not\supset \mathfrak{p}$, $\mathfrak{q}_a \not\supset \mathfrak{p}'$, and $\mathfrak{q}_a \cup \mathfrak{p} \neq \mathfrak{q}_a \cup \mathfrak{p}'$. Now $q \cap (\mathfrak{q}_a \cup \mathfrak{p} \cup \mathfrak{p}') = (q \cap \mathfrak{u}_a) \cap (\mathfrak{q}_a \cup \mathfrak{p} \cup \mathfrak{p}') = \mathfrak{q}_a \cup (q \cap \mathfrak{u}_a \cap (\mathfrak{p} \cup \mathfrak{p}')) = \mathfrak{q}_a \cup (q \cap (\mathfrak{p} \cup \mathfrak{p}'))$ since $\mathfrak{L}_a$ is modular. If $q \cap (\mathfrak{p} \cup \mathfrak{p}') \neq a$, we have $q \cap (\mathfrak{p} \cup \mathfrak{p}') \supset \mathfrak{p}_1 > a$. If $\mathfrak{p}' = \mathfrak{p}_1$, we have $q \supset \mathfrak{p}'$ and hence $\mathfrak{q}_a \supset \mathfrak{p}'$ contrary to hypothesis. Thus $\mathfrak{p}_1 \neq \mathfrak{p}$ and $\mathfrak{p}_1 \neq \mathfrak{p}'$. Now $\mathfrak{p} \cup \mathfrak{p}' \supset \mathfrak{p} \cup \mathfrak{p}_1 \supset \mathfrak{p}$ and $\mathfrak{p} \cup \mathfrak{p}_1 \neq \mathfrak{p}$. Hence $\mathfrak{p} \cup \mathfrak{p}' = \mathfrak{p} \cup \mathfrak{p}_1$ by the Birkhoff condition. Since $q \supset \mathfrak{p}_1$ we have $\mathfrak{q}_a \supset \mathfrak{p}_1$ and hence $\mathfrak{q}_a \cup \mathfrak{p} \supset \mathfrak{p}_1 \cup \mathfrak{p} \supset \mathfrak{p}'$. But then $\mathfrak{q}_a \cup \mathfrak{p} = \mathfrak{q}_a \cup \mathfrak{p}'$ which contradicts the definition of $\mathfrak{p}$ and $\mathfrak{p}'$. Thus $q \cap (\mathfrak{p} \cup \mathfrak{p}') = a$ and $q \cap (\mathfrak{q}_a \cup \mathfrak{p} \cup \mathfrak{p}') = \mathfrak{q}_a$.

Now suppose that $\mathfrak{q}_a$ is *not* principal. Let X be the set of all elements x such that $q \supset x \supset \mathfrak{q}_a$, $q \neq x$. If $x \in X$, let $\mathfrak{p}_x = q \cap (x \cup \mathfrak{p} \cup \mathfrak{p}')$. Clearly X generates $\mathfrak{q}_a$. We shall show

(1) There exists an $x_0 \in X$ such that $x \cup \mathfrak{p} \cup \mathfrak{p}' > \mathfrak{p}_x > x$ for all $x \in X$, $x_0 \supset x$.

(2) The set of ideals $\mathfrak{p}_x$, $x \in X$, $x_0 \supset x$, generates $\mathfrak{q}_a$.

(1) Since $\mathfrak{q}_a \cup \mathfrak{p} \not\supset \mathfrak{p}'$ and X generates $\mathfrak{q}_a$, there exists an element $x_0 \in X$ such that $x_0 \cup \mathfrak{p} \not\supset \mathfrak{p}'$. Let $x_0 \supset x$, $x \in X$ and suppose that $x = \mathfrak{p}_x$. Since x is a proper divisor of a we have $\mathfrak{u}_x > \mathfrak{q}_x$. By the Birkhoff condition $x \cup \mathfrak{p} > x$,

$x\cup\mathfrak{p}'>x$ and hence $x\cup\mathfrak{p}$, $x\cup\mathfrak{p}'$ belong to $\mathfrak{L}_x$. Now $\mathfrak{q}_x\not\supset x\cup\mathfrak{p}\cup\mathfrak{p}'$ since otherwise $q\supset\mathfrak{p}$. Hence by the modularity of $\mathfrak{L}_x$ we have $x\cup\mathfrak{p}\cup\mathfrak{p}'>\mathfrak{q}_x\cap(x\cup\mathfrak{p}\cup\mathfrak{p}')$. Then $x\cup\mathfrak{p}\cup\mathfrak{p}'\supset q\cap(x\cup\mathfrak{p}\cup\mathfrak{p}')\supset\mathfrak{q}_x\cap(x\cup\mathfrak{p}\cup\mathfrak{p}')$ and $q\cap(x\cup\mathfrak{p}\cup\mathfrak{p}')\neq x\cup\mathfrak{p}\cup\mathfrak{p}'$. Thus $x=\mathfrak{p}_x=q\cap(x\cup\mathfrak{p}\cup\mathfrak{p}')=\mathfrak{q}_x\cap(x\cup\mathfrak{p}\cup\mathfrak{p}')$ and hence $x\cup\mathfrak{p}\cup\mathfrak{p}'>x$. But then $x\cup\mathfrak{p}\cup\mathfrak{p}'\supset x\cup\mathfrak{p}\supset x$ and if $x=x\cup\mathfrak{p}$ we have $q\supset\mathfrak{p}$ which is impossible. Thus $x\cup\mathfrak{p}=x\cup\mathfrak{p}\cup\mathfrak{p}'\supset\mathfrak{p}'$ and $x_0\cup\mathfrak{p}\supset x\cup\mathfrak{p}\supset\mathfrak{p}'$ contrary to the definition of x_0. Hence $x\neq\mathfrak{p}_x$. Let $\mathfrak{p}_x\supset\mathfrak{p}_x'>x$. Clearly $x\cup\mathfrak{p}\cup\mathfrak{p}'\supset x\cup\mathfrak{p}\cup\mathfrak{p}_x'\supset x\cup\mathfrak{p}$ and $x\cup\mathfrak{p}\cup\mathfrak{p}'>x\cup\mathfrak{p}$ by the Birkhoff condition. If $x\cup\mathfrak{p}\cup\mathfrak{p}_x'=x\cup\mathfrak{p}$ we have $x\cup\mathfrak{p}\supset\mathfrak{p}_x'$ and $x=q\cap(x\cup\mathfrak{p})\supset\mathfrak{p}_x'$ which contradicts $\mathfrak{p}_x'>x$. Hence $x\cup\mathfrak{p}\cup\mathfrak{p}'=x\cup\mathfrak{p}\cup\mathfrak{p}_x'$. But then $x\cup\mathfrak{p}\cup\mathfrak{p}_x'\supset\mathfrak{p}_x\supset\mathfrak{p}_x'$ and $x\cup\mathfrak{p}\cup\mathfrak{p}_x'>\mathfrak{p}_x'$. If $x\cup\mathfrak{p}\cup\mathfrak{p}_x'=\mathfrak{p}_x$, then $q\supset x\cup\mathfrak{p}\cup\mathfrak{p}'$ which is impossible. Hence $\mathfrak{p}_x=\mathfrak{p}_x'$ and $x\cup\mathfrak{p}\cup\mathfrak{p}'>\mathfrak{p}_x>x$.

(2) Clearly $\mathfrak{p}_x\supset\mathfrak{q}_a$ for every x since $\mathfrak{p}_x\supset x\supset\mathfrak{q}_a$. Now let $a_1\in\mathfrak{q}_a$. Then $a_1\supset\mathfrak{q}_a=q\cap(\mathfrak{q}_a\cup\mathfrak{p}\cup\mathfrak{p}')$ and hence $a_1\supset q\cap(a_2\cup\mathfrak{p}\cup\mathfrak{p}')$ where $a_2\in\mathfrak{q}_a$ by Theorem 2.2. Let $x=x_0\cap a_2$. Then $x\in X$ and $a_1\supset q\cap(x\cup\mathfrak{p}\cup\mathfrak{p}')=\mathfrak{p}_x$. Hence each element of $\mathfrak{q}_a$ divides some $\mathfrak{p}_x$ and thus the ideals $\mathfrak{p}_x$ generate $\mathfrak{q}_a$.

Now let y be an arbitrary element of $\mathfrak{q}_a\cup\mathfrak{p}$. Then $y\supset\mathfrak{q}_a$ and hence $y\supset\mathfrak{p}_x$ where $x_0\supset x$ by (2). But then by (1) $x\cup\mathfrak{p}\cup\mathfrak{p}'>\mathfrak{p}_x>x$ and $\mathfrak{p}_x\not\supset\mathfrak{p}$ since otherwise $q\supset\mathfrak{p}$. Now $x\cup\mathfrak{p}\cup\mathfrak{p}'\supset\mathfrak{p}_x\cup\mathfrak{p}\supset\mathfrak{p}_x$ and $\mathfrak{p}_x\cup\mathfrak{p}\neq\mathfrak{p}_x$. Hence $x\cup\mathfrak{p}\cup\mathfrak{p}'=\mathfrak{p}_x\cup\mathfrak{p}$ which gives $\mathfrak{p}_x\cup\mathfrak{p}\supset\mathfrak{p}'$. Thus $y\cup\mathfrak{p}\supset\mathfrak{p}'$ for every y and hence $\mathfrak{q}_a\cup\mathfrak{p}\supset\mathfrak{p}'$. The assumption that $\mathfrak{q}_a$ is *not* principal has thus led to a contradiction and we conclude that $\mathfrak{q}_a$ is principal, say $\mathfrak{q}_a=(a_1)$. Since a_1 is a proper divisor of a, by hypothesis we have $\mathfrak{u}_{a_1}>\mathfrak{q}_{a_1}$. Hence $\mathfrak{q}_{a_1}\cup(a_1\cup\mathfrak{p}\cup\mathfrak{p}')>\mathfrak{q}_{a_1}$. But $\mathfrak{q}_{a_1}\cap(a_1\cup\mathfrak{p}\cup\mathfrak{p}')=q\cap(a_1\cup\mathfrak{p}\cup\mathfrak{p}')=a_1$ and $a_1\cup\mathfrak{p}\cup\mathfrak{p}'\not>a_1$. Hence $\mathfrak{L}_{a_1}$ is non-modular contrary to assumption. Thus $\mathfrak{q}_a$ is simple and hence $\mathfrak{c}$ is a simple ideal of $\mathfrak{L}_a$.

LEMMA 5.3. *Let $\mathfrak{S}$ be a Birkhoff lattice satisfying the ascending chain condition. Then if $\mathfrak{L}_a$ is modular for every a, $\mathfrak{L}_a$ is archimedean.*

For let $a=q_1\cap\cdots\cap q_n$ be a reduced decomposition of a into irreducibles. Then a has the reduced representation $a=\mathfrak{c}_1\cap\cdots\cap\mathfrak{c}_k$ where $\mathfrak{c}_i$ is a characteristic ideal associated with q_i. By Lemma 5.2, $\mathfrak{c}_i$ is a simple ideal of $\mathfrak{L}_a$. Hence since $\mathfrak{L}_a$ is modular we have $\mathfrak{u}_a>\mathfrak{c}_1>\mathfrak{c}_1\cap\mathfrak{c}_2>\cdots>\mathfrak{c}_1\cap\cdots\cap\mathfrak{c}_k=a$. Thus $\mathfrak{L}_a$ is archimedean of length k.

Lemmas 5.1–5.3 together give Theorem 5.1.

COROLLARY 5.1. *Let $\mathfrak{S}$ be a Birkhoff lattice satisfying the ascending chain condition. Let the number of components in the reduced decompositions of an element a be unique. Then in any two reduced decompositions of a, each component of one decomposition may be replaced by a suitably chosen component of the other.*

For by Lemma 5.1, the two decompositions give two reduced representations of a as a cross-cut of simple ideals of $\mathfrak{L}_a$. However, since $\mathfrak{L}_a$ is modular, B2 is satisfied and the replacement property follows from the dual of Lemma 1.3.

COROLLARY 5.2. *Let $\mathfrak{S}$ be a Birkhoff lattice satisfying the ascending chain condition. Then if the number of components in the decompositions of an element a is unique, that number is simply the length of* $\mathfrak{L}_a$.

COROLLARY 5.3. *Let $\mathfrak{S}$ be a complemented Birkhoff lattice in which every element can be expressed as a cross-cut of a finite number of irreducibles. Then the number of components in the reduced decompositions of the null element z is unique if and only if $\mathfrak{S}$ is a complemented modular lattice of finite dimensions.*

For since $\mathfrak{S}$ is complemented, $\mathfrak{L}_z$ is simply $\mathfrak{L}$, the lattice of ideals.

6. **The Mac Lane exchange axiom.** In order to free condition B1 of the covering properties, Mac Lane (Mac Lane [1]) formulated the following axiom.

E_5. If $a\supset b\supset a\cap c$ and $c\neq a\cap c$, then there exists an element $c_1\neq a\cap c$ such that $c\supset c_1\supset a\cap c$ and $b=a\cap(b\cup c_1)$.

Mac Lane showed that E_5 is equivalent to a transposition property of chains and in case covering elements exist, that is, if $b\supset a$, $b\neq a$, implies b' exists such that $b\supset b'>a$, it reduces to B1. Thus both E_5 and the requirement that each element satisfy the Birkhoff condition in the lattice of ideals are generalizations of B1. We shall be particularly interested in the conditions under which they are equivalent.

THEOREM 6.1. *Every Birkhoff lattice satisfies* E_5.

Proof. Let $a\supset b\supset a\cap c$ and $c\neq a\cap c$. Then by Theorem 2.1 and ideal $\mathfrak{p}$ exists such that $c\supset\mathfrak{p}>a\cap b$. Now $b\not\supset\mathfrak{p}$ since otherwise $a\cap c\supset\mathfrak{p}$ which is impossible. Hence $b\cup\mathfrak{p}>b$ by the Birkhoff condition. But then $b\cup\mathfrak{p}\supset a\cap(b\cup\mathfrak{p})$ $\supset b$ and if $b\cup\mathfrak{p}=a\cap(b\cup\mathfrak{p})$ we have $a\supset\mathfrak{p}$ which is impossible. Hence $b=a\cap(b\cup\mathfrak{p})$. Thus by Theorem 2.2 an element $p\in\mathfrak{p}$ exists such that $b=a\cap(b\cup p)$. Let $c_1=c\cap p$. Then $c\supset c_1\supset\mathfrak{p}>a\cap c$ and hence $c\supset c_1\supset a\cap c$, $c_1\neq a\cap c$. Also $b=a\cap(b\cup p)\supset a\cap(b\cup c_1)\supset b$. Thus $b=a\cap(b\cup c_1)$ and c_1 satisfies the requirements of E_5.

THEOREM 6.2. *Let $\mathfrak{S}$ satisfy E_5 and have the property that each element is covered by only a finite number of covering ideals. Then $\mathfrak{S}$ is a Birkhoff lattice.*

Proof. Let $\mathfrak{a}\supset a$, $\mathfrak{p}>a$ and $\mathfrak{a}\not\supset\mathfrak{p}$. Let $\mathfrak{p}, \mathfrak{p}_1, \cdots, \mathfrak{p}_n$ be the finite number of ideals covering a. Now if $\mathfrak{p}\cup\mathfrak{a}\not>\mathfrak{a}$, we have $\mathfrak{p}\cup\mathfrak{a}\supset\mathfrak{c}\supset\mathfrak{a}$, $\mathfrak{p}\cup\mathfrak{a}\neq\mathfrak{c}\neq\mathfrak{a}$. Since $\mathfrak{p}\supset\mathfrak{p}\cap\mathfrak{c}\supset a$ and $\mathfrak{c}\not\supset\mathfrak{p}$ we have $\mathfrak{p}\cap\mathfrak{c}=a$ and hence by Theorem 2.2 elements $p'\in\mathfrak{p}$ and $c\in\mathfrak{c}$ exist such that $p'\cap c=a$. Since $\mathfrak{c}\neq\mathfrak{a}$, there exists an element $b'\in\mathfrak{a}$ such that $b'\not\supset\mathfrak{c}$. Let $b=c\cap b'$. Then $b\not\supset\mathfrak{c}$ and $c\supset b\supset c\cap p'$. Now since $\mathfrak{p}\not\supset\mathfrak{p}_i$ $(i=1, \cdots, n)$ elements p_i' exist such that $p_i'\in\mathfrak{p}$ and $p_i'\not\supset\mathfrak{p}_i$ $(i=1, \cdots, n)$. Set $p=p'\cap p_1'\cap\cdots\cap p_n'$. Then $p\in\mathfrak{p}$, $p'\supset p$ and $p\not\supset\mathfrak{p}_i$ $(i=1, \cdots, n)$. Clearly $a=c\cap p'\supset c\cap p\supset a$ implies $c\cap p=a$. Hence $c\supset b\supset c\cap p$ and $p\neq c\cap p$. Thus by E_5 an element p_1 exists such that $p_1\neq a$, $p\supset p_1\supset a$ and

$b=c\cap(b\cup p_1)$. Now $p_1\supset\mathfrak{p}'>a$ and $\mathfrak{p}'\neq\mathfrak{p}_i$ $(i=1,\cdots,n)$ since otherwise $p\supset p_1\supset\mathfrak{p}_i$ contrary to the definition of p. Hence $\mathfrak{p}'=\mathfrak{p}$. But then $b=c\cap(b\cup p_1)\supset\mathfrak{c}\cap(\mathfrak{a}\cup\mathfrak{p})\supset\mathfrak{c}$ which contradicts $b\not\supset\mathfrak{c}$. Hence $\mathfrak{p}\cup\mathfrak{a}>\mathfrak{a}$ and $\mathfrak{S}$ is a Birkhoff lattice.

Now by Lemma 4.6, if $\mathfrak{S}$ is a Birkhoff lattice in which the ascending chain condition holds and every three ideals covering a principal ideal generate a Boolean algebra, then each a is covered by only a finite number of ideals. However, Theorem 6.2 does not enable us to replace the Birkhoff condition in the lemma by E_5 since the proof of the finiteness required the Birkhoff condition. To carry out this replacement we first replace the condition that every three ideals covering a principal ideal generate a Boolean algebra by an equivalent condition.

THEOREM 6.3. *Let $\mathfrak{S}$ be a Birkhoff lattice satisfying the ascending chain condition. Then every three ideals covering a principal ideal generate a Boolean algebra if and only if $a\cup b\supset\mathfrak{q}>a\cap b$ implies $a\supset\mathfrak{q}$ or $b\supset\mathfrak{q}$.*

Proof. Let every three ideals covering a principal ideal generate a Boolean algebra and suppose that $a\cup b\supset\mathfrak{q}>a\cap b$ but $a\not\supset\mathfrak{q}$, $b\not\supset\mathfrak{q}$. Then $a, b\neq a\cap b$. For if $a=a\cap b$, then $b=a\cup b\supset\mathfrak{q}$ contrary to assumption. Now with b fixed let a be maximal such that $a\cup b\supset\mathfrak{q}>a\cap b$ and $a\not\supset\mathfrak{q}$, $b\not\supset\mathfrak{q}$ for some $\mathfrak{q}$. Let $b\supset\mathfrak{p}>a\cap b$. Suppose $\mathfrak{a}=\mathfrak{p}\cup a\not\supset\mathfrak{q}$. Then $\mathfrak{a}\cup b=a\cup\mathfrak{p}\cup b=a\cup b\supset\mathfrak{q}$ and $\mathfrak{a}\cap b\supset a\cap b$. Now $\mathfrak{a}\neq a$ since otherwise $a\supset\mathfrak{p}$ and $a\cap b\supset\mathfrak{p}>a\cap b$ which is impossible. Let a_1' be an element of $\mathfrak{a}$ such that $a_1'\not\supset\mathfrak{q}$. Now $a\cup b\supset b\supset\mathfrak{p}$ and hence $a\cup b\supset\mathfrak{p}\cup a=\mathfrak{a}$. Thus $a_1=(a\cup b)\cap a_1'\supset\mathfrak{a}$ and $a\cup b\supset a_1$. But $a\cup b\supset a_1\cup b\supset a\cup b$. Hence $a_1\cup b=a\cup b\supset\mathfrak{q}$. Also $a_1\not\supset\mathfrak{q}$ since otherwise $a_1'\supset\mathfrak{q}$. Now $\mathfrak{q}\supset(a_1\cap b)\cap\mathfrak{q}\supset a\cap b$ and $\mathfrak{q}\neq(a_1\cap b)\cap\mathfrak{q}$ since otherwise $b\supset\mathfrak{q}$. Hence $\mathfrak{q}>a\cap b=(a_1\cap b)\cap\mathfrak{q}$. By the Birkhoff condition we have $\mathfrak{q}_1=\mathfrak{q}\cup(a_1\cap b)>a_1\cap b$. Since $a_1\cup b\supset\mathfrak{q}$ we have $a_1\cup b\supset\mathfrak{q}_1>a_1\cap b$. Clearly $a_1\not\supset\mathfrak{q}_1$, $b\not\supset\mathfrak{q}_1$, and $a_1\neq a$. This contradicts the maximal property of a. Hence we have $\mathfrak{p}\cup a\supset\mathfrak{q}$.

Now let $a\supset\mathfrak{p}_1>a\cap b$. $\mathfrak{p}_1\neq\mathfrak{q}$ since otherwise $a\supset\mathfrak{q}$ and $\mathfrak{p}_1\neq\mathfrak{p}$ since otherwise $a\cap b\supset\mathfrak{p}>a\cap b$. Hence $\mathfrak{p}_1\cup\mathfrak{p}\not\supset\mathfrak{q}$ since every three ideals covering $a\cap b$ generate a Boolean algebra. Thus $p_1\cup\mathfrak{p}\not\supset\mathfrak{q}$ for some $p_1\in\mathfrak{p}_1$. Set $x=p_1\cap a$. Then $a\supset x\supset a\cap b$ and $x\neq a\cap b$, $x\cup\mathfrak{p}\not\supset\mathfrak{q}$. Let a_2 be a maximal such x. Then $a_2\neq a$ since $a\cup\mathfrak{p}\supset\mathfrak{q}$. Hence $a\supset\mathfrak{p}_2>a_2$ for some ideal $\mathfrak{p}_2$. Then if $\mathfrak{p}_2\cup\mathfrak{p}\not\supset\mathfrak{q}$ we have $p_2\cup\mathfrak{p}\not\supset\mathfrak{q}$ for some $p_2\in\mathfrak{p}_2$. Let $a_2'=a\cap p_2$. Then $a\supset a_2'\supset a\cap b$, $a_2'\neq a\cap b$ and $a_2'\cup\mathfrak{p}\not\supset\mathfrak{q}$ contrary to the maximal property of a_2. Hence $\mathfrak{p}_2\cup\mathfrak{p}\supset\mathfrak{q}$. Let $\mathfrak{q}_2=a_2\cup\mathfrak{q}$ and $\mathfrak{p}_3=a_2\cup\mathfrak{p}$. We have $a_2\not\supset\mathfrak{p}$ since $a\not\supset\mathfrak{p}$ and hence $\mathfrak{p}_3$, $\mathfrak{q}_2$, $\mathfrak{p}_2>a_2$ by the Birkhoff condition. Clearly $\mathfrak{p}_2\cup\mathfrak{p}_3=\mathfrak{p}_2\cup a_2\cup\mathfrak{p}\supset a_2\cup\mathfrak{q}=\mathfrak{q}_2$. Now $\mathfrak{p}_2\neq\mathfrak{q}_2$ since otherwise $a\supset\mathfrak{q}$. Also $\mathfrak{q}_2\neq\mathfrak{p}_3$ since otherwise $a_2\cup\mathfrak{p}\supset\mathfrak{q}$ contrary to the definition of a_2. Finally $\mathfrak{p}_2\neq\mathfrak{p}_3$ since $a\not\supset\mathfrak{p}$. Thus $\mathfrak{p}_2$, $\mathfrak{q}_2$, $\mathfrak{p}_3$ do not generate a Boolean algebra, which contradicts our hypothesis. Hence either $a\supset\mathfrak{q}$ or $b\supset\mathfrak{q}$.

On the other hand if three ideals $\mathfrak{a}$, $\mathfrak{b}$, $\mathfrak{c}$ covering d do not generate a Boolean algebra, then $\mathfrak{a}\cup\mathfrak{b}\supset\mathfrak{c}$ say. Since $d=\mathfrak{a}\cap\mathfrak{b}$, elements $a\in\mathfrak{a}$ and $b\in\mathfrak{b}$

exist such that $d=a\cap b$. Hence $a\cup b\supset\mathfrak{a}\cup\mathfrak{b}\supset\mathfrak{c}>a\cap b=d$. If $a\supset\mathfrak{c}$, then $a\supset\mathfrak{a}\cup\mathfrak{c}\supset\mathfrak{b}$ and $d=a\cap b\supset\mathfrak{b}$ which is impossible. Hence $a\not\supset\mathfrak{c}$ and $b\not\supset\mathfrak{c}$. Thus $a\cup b\supset\mathfrak{c}>a\cap b$ but $a\not\supset\mathfrak{c}$ and $b\not\supset\mathfrak{c}$. This completes the proof of the theorem.

We show now that if the ascending chain condition holds and the condition of Theorem 6.3 is satisfied, then E_6 is equivalent to the Birkhoff condition. A preliminary lemma is required.

LEMMA 6.1. *$\mathfrak{S}$ is a Birkhoff lattice if and only if $b\supset a$, $b\not\supset\mathfrak{p}>a$ implies $\mathfrak{p}\cup b>b$.*

The necessity of the condition is obvious. To prove the sufficiency let $\mathfrak{b}>a$, $\mathfrak{c}\supset a$ and $\mathfrak{c}\not\supset\mathfrak{b}$. If $\mathfrak{b}\cup\mathfrak{c}\not>\mathfrak{c}$ we have $\mathfrak{b}\cup\mathfrak{c}\supset\mathfrak{d}\supset\mathfrak{c}$ where $\mathfrak{b}\cup\mathfrak{c}\neq\mathfrak{d}\neq\mathfrak{c}$. Since $\mathfrak{d}\not\supset\mathfrak{b}$ an element $d\in\mathfrak{d}$ exists such that $d\not\supset\mathfrak{b}$. Since $d\supset\mathfrak{c}$ and $\mathfrak{c}\not\supset\mathfrak{d}$ there exists an element $c\in\mathfrak{c}$ such that $d\supset c$ and $c\not\supset\mathfrak{d}$. Now $c\supset a$, $c\not\supset\mathfrak{b}>a$. Hence $c\cup\mathfrak{b}>c$. But $c\cup\mathfrak{b}\supset d\cap(c\cup\mathfrak{b})\supset c$ and $c\cup\mathfrak{b}\neq d\cap(c\cup\mathfrak{b})$ since $d\not\supset\mathfrak{b}$. Thus $c=d\cap(c\cup\mathfrak{b})\supset\mathfrak{d}\cap(\mathfrak{c}\cup\mathfrak{b})=\mathfrak{d}$ which contradicts $c\not\supset\mathfrak{d}$. Hence $\mathfrak{b}\cup\mathfrak{c}>\mathfrak{c}$ and $\mathfrak{S}$ is a Birkhoff lattice.

THEOREM 6.4. *Let $\mathfrak{S}$ satisfy* E_6, *the ascending chain condition, and let $a\cup b\supset\mathfrak{q}>a\cap b$ imply $a\supset\mathfrak{q}$ or $b\supset\mathfrak{q}$. Then $\mathfrak{S}$ is a Birkhoff lattice.*

Proof. Let X be the set of all elements x such that $y\supset x$, $\mathfrak{p}>x$, $y\not\supset\mathfrak{p}$, and $\mathfrak{p}\cup y\not>y$ for some y and $\mathfrak{p}$. If $\mathfrak{S}$ is not a Birkhoff lattice, then X is non-empty by Lemma 6.1. Let a be a maximal element of X. Then b and $\mathfrak{p}$ exist such that $b\supset a$, $b\not\supset\mathfrak{p}>a$ and $b\cup\mathfrak{p}\supset\mathfrak{c}>b$, $b\cup\mathfrak{p}\neq\mathfrak{c}$. Now $\mathfrak{p}\supset\mathfrak{p}\cap\mathfrak{c}\supset a$ and $\mathfrak{p}\neq\mathfrak{p}\cap\mathfrak{c}$ since otherwise $b\cup\mathfrak{p}=\mathfrak{c}$. Hence $\mathfrak{p}\cap\mathfrak{c}=a$. Let $p\in\mathfrak{p}$, $c\in\mathfrak{c}$ such that $p\cap c=a$. Then $p\cap c=p\cap b=a$ and $c\neq b$. Hence by E_6 an element p_1 exists such that $p\supset p_1\supset a$, $p_1\neq a$ and $b=c\cap(b\cup p_1)$. Now $b\cup p_1\not\supset\mathfrak{c}$ since otherwise $b=c\cap(b\cup p_1)\supset\mathfrak{c}$ which contradicts $\mathfrak{c}>b$. Now suppose that we have found $p_1,\cdots,p_k$ such that $b\cup p_1\cup p_2\cup\cdots\cup p_k\not\supset\mathfrak{c}$ and $(b\cup p_1\cup\cdots\cup p_i)\cap p_{i+1}=a$ $(i=1,\cdots,k-1)$. Since $b\cup p_1\cup p_2\cup\cdots\cup p_k\not\supset\mathfrak{c}$ we have $b\cup p_1\cup\cdots\cup p_k\not\supset\mathfrak{p}$ and hence $(b\cup p_1\cup\cdots\cup p_k)\cap\mathfrak{p}=a$. Thus $p'_{k+1}\in\mathfrak{p}$ exists such that $(b\cup p_1\cup\cdots\cup p_k)\cap p'_{k+1}=a$. Let $p''_{k+1}=p'_{k+1}\cap p$. Then $(b\cup p_1\cup\cdots\cup p_k)\cap p''_{k+1}=a$ and $c\cap p_{k+1}=b\cap p''_{k+1}=a$, $p''_{k+1}\neq a$. Hence by E_6 an element p_{k+1} exists such that $p''_{k+1}\supset p_{k+1}\supset a$, $p_{k+1}\neq a$ and $b=c\cap(b\cup p_{k+1})$. Then $b\cup p_{k+1}\not\supset\mathfrak{c}$ since otherwise $b=c\cap(b\cup p_{k+1})\supset\mathfrak{c}$. Now $\mathfrak{c}\supset\mathfrak{c}\cap((b\cup p_1\cup\cdots\cup p_k)\cap(b\cup p_{k+1}))\supset b$ and $\mathfrak{c}\neq\mathfrak{c}\cap((b\cup p_1\cup\cdots\cup p_k)\cap(b\cup p_{k+1}))$ since otherwise $b\cup p_{k+1}\supset(b\cup p_1\cup\cdots\cup p_k)\cap(b\cup p_{k+1})\supset\mathfrak{c}$. Hence since $\mathfrak{c}>b$ we have $\mathfrak{c}>\mathfrak{c}\cap((b\cup p_1\cup\cdots\cup p_k)\cap(b\cup p_{k+1}))=b$. But now since b is a proper divisor of a, by the maximal property of a we must have $\mathfrak{c}\cup((b\cup p_1\cup\cdots\cup p_k)\cap(b\cup p_{k+1}))>(b\cup p_1\cup\cdots\cup p_k)\cap(b\cup p_{k+1})$. Now suppose that $b\cup p_1\cup\cdots\cup p_{k+1}\supset\mathfrak{c}$. Then $(b\cup p_1\cup\cdots\cup p_k)\cup(b\cup p_{k+1})\supset\mathfrak{c}\cup((b\cup p_1\cup\cdots\cup p_k)\cap(b\cup p_{k+1}))>(b\cup p_1\cup\cdots\cup p_k)\cap(b\cup p_{k+1})$. Hence by hypothesis either $b\cup p_1\cup\cdots\cup p_k\supset\mathfrak{c}$ or $b\cup p_{k+1}\supset\mathfrak{c}$ both of which are impossible. Thus $b\cup p_1\cup\cdots\cup p_{k+1}\not\supset\mathfrak{c}$. By induction we get an infinite chain $b\subset b\cup p_1\subset b\cup p_1\cup p_2\subset\cdots\subset b\cup p_1\cup\cdots$

$\cup p_i \subset \cdots$ and $b\cup p_1 \cup \cdots \cup p_i \neq b\cup p_1 \cup \cdots \cup p_{i+1}$ since $(b\cup p_1\cup \cdots \cup p_i)\cap p_{i+1}=a$. This chain contradicts the ascending chain condition and hence $\mathfrak{S}$ is a Birkhoff lattice.

The condition $a\cup b\supset \mathfrak{q} > a\cap b$ implies $a\supset \mathfrak{q}$ or $b\supset \mathfrak{q}$, may be given a purely combinatorial statement as follows:

THEOREM 6.5. *$a\cup b\supset \mathfrak{q} > a\cap b$ implies $a\supset \mathfrak{q}$ or $b\supset \mathfrak{q}$ if and only if*
(A) *$a\cup b\supset x\supset a\cap b$, $a\cap x=b\cap x=a\cap b$ implies $x=a\cap b$.*

Proof. Let $a\cup b\supset x\supset a\cap b$, $a\cap x=b\cap x=a\cap b$. If $x\neq a\cap b$, let $x\supset \mathfrak{q} > a\cap b$. Then $a\cup b\supset \mathfrak{q} > a\cap b$ and hence $a\supset \mathfrak{q}$ say. But then $a\cap b=a\cap x\supset \mathfrak{q} > a\cap b$ which is impossible. Hence $x=a\cap b$.

On the other hand, let $a\cup b\supset \mathfrak{q} > a\cap b$. If $a \not\supset \mathfrak{q}$ and $b\not\supset \mathfrak{q}$ we have $a\cap \mathfrak{q} =b\cap \mathfrak{q}=a\cap b$. Hence for some $x\in \mathfrak{q}$ we have $a\cap x=b\cap x=a\cap b$ and $a\cup b\supset x$ by Theorem 2.2. But then $a\cup b\supset x\supset a\cap b$, $a\cap x=b\cap x=a\cap b$ and $x\neq a\cap b$ which contradicts condition A.

Lemma 4.6 with Theorems 6.1–6.5 give

THEOREM 6.6. *Let $\mathfrak{S}$ satisfy the ascending chain condition. Then every element of $\mathfrak{S}$ is uniquely expressible as a reduced cross-cut of irreducibles if and only if conditions* E_6 *and* A *are satisfied.*

Theorem 6.6 has the following interesting corollary:

COROLLARY 6.1: *Let $\mathfrak{S}$ satisfy the ascending chain condition and let each element of $\mathfrak{S}$ have a unique reduced decomposition into irreducibles. Then a sublattice $\mathfrak{S}'$ of $\mathfrak{S}$ has unique irreducible decompositions if and only if* E_6 *holds in $\mathfrak{S}'$.*

Axiom A is clearly a slightly stronger form of the requirement that every modular sublattice be distributive. In D1 it was shown that under the assumption of both the ascending and descending chain conditions, this weaker condition and B1 were necessary and sufficient for unique decomposition into irreducibles. But A *cannot* be replaced by the requirement that every modular sublattice be distributive in Theorem 6.6 as the example of Figure 1 shows.

The non-principal ideals of $\mathfrak{S}$ are the ideals $\mathfrak{a}$, generated by $a_1, a_2, a_3, \cdots$; $\mathfrak{b}$, generated by $b_1, b_2, b_3, \cdots$; and $\mathfrak{c}$, generated by $c_1, c_2, c_3, \cdots$. Clearly $\mathfrak{a}>z$, $\mathfrak{b}>b$, and $\mathfrak{c}>z$. Now $b\cup a_{2n}=d_n>a_{2n}$ and $b\cup a_{2n+1}=b_{2n+1}>a_{2n+1}$. Hence $b\cup a_i >a_i$. Similarly $b\cup c_i>c_i$. Now let x be any element of $\mathfrak{S}$ not equal to b or z. Then b_1/x is an archimedean lattice and B1 is readily verified in b_1/x since each element has at most two covering elements. Thus we have only to verify the Birkhoff condition for non-principal ideals. Clearly $\mathfrak{b}>\mathfrak{a}$, b, $\mathfrak{c}>z$. Hence $\mathfrak{a}\cup b>b$, $\mathfrak{a}$; $\mathfrak{a}\cup \mathfrak{c}>\mathfrak{a}$, $\mathfrak{c}$; $b\cup \mathfrak{c}>b$, $\mathfrak{c}$. $\mathfrak{a}\cup c_{2n}=e_n>c_{2n}$ and $\mathfrak{a}\cup c_{2n+1}=b_{2n+2}>c_{2n+1}$. Hence $\mathfrak{a}\cup c_i>c_i$. Similarly $\mathfrak{c}\cup a_i>a_i$. Thus every element of $\mathfrak{S}$ satisfies the Birkhoff condition in the lattice of ideals and hence $\mathfrak{S}$ is a Birkhoff lattice. By Theorem 6.1, E_6 holds in $\mathfrak{S}$. Now if $\mathfrak{S}$ contains a modular, non-distribu-

tive sublattice it also contains one of the form $\{u, v, w, x, y\}$ where $v \cup w = w \cup x = v \cup x = u$ and $v \cap w = w \cap x = v \cap x = y$. Since every element not equal to b or z is covered by at most two elements we must have $y = z$. But then $v = a_i$, $w = b$, $x = c_j$ and $v \cup w = a_i \cup b \neq b \cup c_j = w \cup x$ which contradicts $v \cup w$

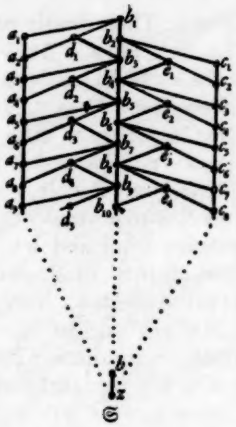


FIG. 1

$= x \cup w$. Hence every modular sublattice of $\mathfrak{S}$ is distributive. However $\mathfrak{S}$ does *not* have unique irreducible decompositions since $z = a_1 \cap b = b \cap c_1 = a_1 \cap c_1$ and a_1, b, c are irreducibles of $\mathfrak{S}$. Axiom A does not hold since $a_i \cup c_j \supset b \supset a_i \cap c_j$ and $a_i \cap b = c_j \cap b = a_i \cap c_j$ but $b \neq a_i \cap c_j = z$.

Since $\mathfrak{a}$, b, $\mathfrak{c}$ generate a modular lattice, $\mathfrak{L}_x$ is modular for every x. Hence by Theorem 5.1, the number of components in the reduced decompositions of each element must be unique. This can be readily verified.

According to Theorem 6.2, if every element of a lattice $\mathfrak{S}$ is covered by only a finite number of ideals, then E_5 implies that $\mathfrak{S}$ is a Birkhoff lattice. We prove now an even stronger theorem, namely, under this restriction E_5 implies that B1 holds in the lattice of ideals. We begin with necessary lemmas.

LEMMA 6.2. B1 *holds in the lattice of ideals of* $\mathfrak{S}$ *if and only if* $\mathfrak{a} > x \cap \mathfrak{a}$ *implies* $x \cup \mathfrak{a} > x$ *for every* $\mathfrak{a} \in \mathfrak{L}$ *and* $x \in \mathfrak{S}$.

For if B1 holds in $\mathfrak{L}$, then clearly $\mathfrak{a} > x \cap \mathfrak{a}$ implies $x \cup \mathfrak{a} > x$. Now let $\mathfrak{a} > x \cap \mathfrak{a}$ imply $\mathfrak{a} \cup x > x$ for each $\mathfrak{a}$ and x. Suppose that B1 does not hold in $\mathfrak{L}$. Then ideals $\mathfrak{a}$ and $\mathfrak{b}$ exist such that $\mathfrak{a} > \mathfrak{a} \cap \mathfrak{b}$ but $\mathfrak{a} \cup \mathfrak{b} \supset \mathfrak{c} \supset \mathfrak{b}$, $\mathfrak{a} \cup \mathfrak{b} \neq \mathfrak{c} \neq \mathfrak{b}$. Let $x_1 \supset \mathfrak{b}$, $x_1 \not\supset \mathfrak{a}$. Such an x_1 always exists since $\mathfrak{b} \not\supset \mathfrak{a}$. Also since $\mathfrak{b} \not\supset \mathfrak{c}$, an element x_2 exists such that $x_2 \supset \mathfrak{b}$, $x_2 \not\supset \mathfrak{c}$. Finally since $\mathfrak{b} \cup \mathfrak{c} \not\supset \mathfrak{a}$ there is an element x_3 such that $x_3 \supset \mathfrak{b}$, $x_3 \cup \mathfrak{c} \not\supset \mathfrak{a}$. Let $x = x_1 \cap x_2 \cap x_3$. Then $x \supset \mathfrak{b}$, $x \not\supset \mathfrak{a}$, and $x \cup \mathfrak{c} \not\supset \mathfrak{a}$. Now $\mathfrak{a} \supset \mathfrak{a} \cap x \supset \mathfrak{a} \cap \mathfrak{b}$ and $\mathfrak{a} \neq \mathfrak{a} \cap x$. Hence $\mathfrak{a} \cap x = \mathfrak{a} \cap \mathfrak{b}$ and thus $\mathfrak{a} > \mathfrak{a} \cap x$. By

hypothesis then $x\cup\mathfrak{a}>x$. Now $x\cup\mathfrak{a}\supset\mathfrak{c}\cup x\supset x$ and $\mathfrak{c}\cup x\neq x$. Hence $x\cup\mathfrak{a}=\mathfrak{c}\cup x$ which implies $x\cup\mathfrak{c}\supset\mathfrak{a}$ contrary to the definition of x. Hence B1 holds in $\mathfrak{L}$.

LEMMA 6.3. *Let $\mathfrak{S}$ be a Birkhoff lattice. Then if $\mathfrak{a}>x\cap\mathfrak{a}$ and $x\cup\mathfrak{a}\not>x$, each $x\cap a$, $a\in\mathfrak{a}$ is covered by an infinite number of ideals.*

For let $\mathfrak{a}>x\cap\mathfrak{a}$ and $x\cup\mathfrak{a}\not>x$. Then clearly $a\cap x\neq a$ for every $a\in\mathfrak{a}$ since otherwise $x\supset a\supset\mathfrak{a}$ and $\mathfrak{a}\not>x\cap\mathfrak{a}$. Hence $a\supset\mathfrak{p}_a>a\cap x$ for some ideal $\mathfrak{p}_a$ by Theorem 2.1. Let S_a denote the set of all ideals $\mathfrak{p}_a$. Now $x\not\supset\mathfrak{p}_a$, since otherwise $a\cap x\supset\mathfrak{p}_a>a\cap x$ which is impossible. Thus $\mathfrak{p}_a\supset x\cap\mathfrak{p}_a\supset a\cap x$ and $\mathfrak{p}_a\neq x\cap\mathfrak{p}_a$. Hence $x\cap\mathfrak{p}_a=a\cap x$ and $\mathfrak{p}_a>x\cap\mathfrak{p}_a$ where $x\cap\mathfrak{p}_a$ is a principal ideal of $\mathfrak{L}$. Since $\mathfrak{S}$ is a Birkhoff lattice we have $x\cup\mathfrak{p}_a>x$ for every $\mathfrak{p}_a$.

Now in S_a we set $\mathfrak{p}_a\sim\mathfrak{p}_a'$ if and only if $x\cup\mathfrak{p}_a=x\cup\mathfrak{p}_a'$. Then $\sim$ is an equivalence relation which separates S_a into mutually exclusive sets of ideals. Let B_a denote an arbitrary equivalence class and let $\mathfrak{b}_a=\Sigma(B_a)$. If $\mathfrak{p}_a\in B_a$, then $x\cup\mathfrak{p}_a=x\cup\mathfrak{p}_a'$ for every other ideal $\mathfrak{p}_a'$ of B_a and hence $x\cup\mathfrak{p}_a=x\cup\mathfrak{b}_a$. Thus $x\cup\mathfrak{b}_a>x$. Let T_a denote the set of ideals $\mathfrak{b}_a$. Now if $a\supset a_1\supset\mathfrak{a}$ and $\mathfrak{b}_{a_1}\in T_{a_1}$, let $\mathfrak{b}_{a_1}\supset\mathfrak{p}_{a_1}>x\cap a_1$. Then $\mathfrak{p}_{a_1}\supset(x\cap a)\cap\mathfrak{p}_{a_1}\supset x\cap a_1$ and $\mathfrak{p}_{a_1}\neq(x\cap a)\cap\mathfrak{p}_{a_1}$ since otherwise $x\supset x\cap a\supset\mathfrak{p}_{a_1}$ which contradicts $x\not\supset\mathfrak{p}_{a_1}$. Hence $\mathfrak{p}_{a_1}>(x\cap a)\cap\mathfrak{p}_{a_1}=x\cap a_1$ and $(x\cap a)\cup\mathfrak{p}_{a_1}>x\cap a$ by the Birkhoff condition. Also $a\supset a\cup a_1\supset(x\cap a)\cup\mathfrak{p}_{a_1}>x\cap a$ and hence $\mathfrak{p}_a=(x\cap a)\cup\mathfrak{p}_{a_1}$ belongs to S_a. Now $x\cup\mathfrak{p}_a=x\cup(x\cap a)\cup\mathfrak{p}_{a_1}=x\cup\mathfrak{p}_{a_1}$. Let $\mathfrak{b}_{a_1}\supset\mathfrak{p}_{a_1}'>x\cap a_1$. Then $x\cup\mathfrak{p}_a'=x\cup\mathfrak{p}_{a_1}'$ where $\mathfrak{p}_a'=(x\cap a)\cup\mathfrak{p}_{a_1}'$. Thus $x\cup\mathfrak{p}_a=x\cup\mathfrak{p}_{a_1}=x\cup\mathfrak{p}_{a_1}'=x\cup\mathfrak{p}_a'$ and $\mathfrak{p}_a\sim\mathfrak{p}_a'$ in S_a. Hence $\mathfrak{b}_{a_1}\subset\mathfrak{b}_a$ where $\mathfrak{b}_a$ is an ideal of T_a. Now suppose that T_{a_1} contains a second ideal $\mathfrak{b}_{a_1}'$, which is divisible by $\mathfrak{b}_a$. Let $\mathfrak{b}_{a_1}'\supset\mathfrak{p}_{a_1}''>x\cap a_1$. Then $x\cup\mathfrak{p}_{a_1}=x\cup\mathfrak{p}_a=x\cup\mathfrak{b}_a\supset x\cup\mathfrak{b}_{a_1}'\supset x\cup\mathfrak{p}_{a_1}''\supset x$. Since $x\cup\mathfrak{p}_{a_1}''\neq x$, we have $x\cup\mathfrak{p}_{a_1}=x\cup\mathfrak{p}_{a_1}''$ and $\mathfrak{p}_{a_1}\sim\mathfrak{p}_{a_1}''$ in S_{a_1} contrary to assumption. Hence $\mathfrak{b}_{a_1}$ is the only ideal of T_{a_1} divisible by $\mathfrak{b}_a$. Next suppose that T_a contains another ideal $\mathfrak{b}_a'$ such that $\mathfrak{b}_a'\supset\mathfrak{b}_{a_1}$. Then $x\cup\mathfrak{b}_a'\supset x\cup\mathfrak{b}_{a_1}=x\cup\mathfrak{b}_a$ and hence $\mathfrak{b}_a=\mathfrak{b}_a'$. We thus conclude that *each ideal $\mathfrak{b}_{a_1}$ of T_{a_1} is divisible by exactly one ideal $\mathfrak{b}_a$ of T_a and $\mathfrak{b}_{a_1}$ is the only ideal of T_{a_1} which is divisible by $\mathfrak{b}_a$.*

Let n_a denote the cardinal number of the set T_a. If $x\cap a$ is covered by only a finite number of ideals for some a, then n_a is finite for some a and hence has a minimal value for some a_0. If $a_0\supset a\supset\mathfrak{a}$, then $n_a\leqq n_{a_0}$ since distinct ideals of T_a are divisible by distinct ideals of T_{a_0}. But since n_{a_0} is minimal we have $n_a=n_{a_0}$. Hence each ideal of T_{a_0} divides exactly one ideal of T_a. Let $\mathfrak{b}_0$ be an ideal of T_{a_0} and let $\mathfrak{b}_a$ denote the ideal of T_a divisible by $\mathfrak{b}_0$. In general, for any $a\in\mathfrak{a}$, let $\mathfrak{b}_a$ be the ideal of T_a divisible by $x\cup\mathfrak{b}_0$. Such an ideal $\mathfrak{b}_a$ always exists since $\mathfrak{b}_0\supset\mathfrak{b}_a'$ where $\mathfrak{b}_a'\in T_{a\cap a_0}$ and hence $\mathfrak{b}_a\in T_a$ exists such that $\mathfrak{b}_a\supset\mathfrak{b}_a'$. Clearly $x\cup\mathfrak{b}_0=x\cup\mathfrak{b}_a'=x\cup\mathfrak{b}_a\supset\mathfrak{b}_a$ and $\mathfrak{b}_a$ is unique as shown above. If $a\supset a'$, we have $\mathfrak{b}_a\supset\mathfrak{b}_{a'}$ since $x\cup\mathfrak{b}_a=x\cup\mathfrak{b}_{a'}$. Hence $\mathfrak{b}_{a_1}\cap\mathfrak{b}_{a_2}\cap\cdots\cap\mathfrak{b}_{a_n}\supset\mathfrak{b}_{a_1\cap a_2\cap\cdots\cap a_n}$.

Now let $\mathfrak{a}_1=\prod_{a\in\mathfrak{a}}(\mathfrak{b}_a)$. Since $a\supset\mathfrak{b}_a\supset\mathfrak{a}_1$, we have $\mathfrak{a}\supset\mathfrak{a}_1\supset x\cap\mathfrak{a}$. If $\mathfrak{a}_1=x\cap\mathfrak{a}_1$ then $x\cap a$ divides the cross-cut of a finite number of the ideals $\mathfrak{b}_a$ and hence

$x\cap a\supset \mathfrak{b}_{a'}$ for some a'. But then $(x\cap a)\cap a'\supset \mathfrak{b}_{a'}\cap \mathfrak{b}_{a'}\supset \mathfrak{b}_{a'}\supset \mathfrak{b}_{a\cap a'}\supset x\cap(a\cap a')$ and $\mathfrak{b}_{a\cap a'}=x\cap(a\cap a')$ contrary to the definition of $\mathfrak{b}_{a\cap a'}$. Thus $\mathfrak{a}_1\neq x\cap \mathfrak{a}$ and hence $\mathfrak{a}=\mathfrak{a}_1$ since $\mathfrak{a}>x\cap\mathfrak{a}$. But then $x\cup \mathfrak{b}_0\supset x\cup\mathfrak{a}_1\supset x\cup \mathfrak{a}\supset x$. Since $x\cup\mathfrak{b}_0>x$ and $x\not\supset\mathfrak{a}$ we have $x\cup\mathfrak{a}=x\cup\mathfrak{b}_0>x$ which contradicts $x\cup\mathfrak{a}\not> x$. Hence each

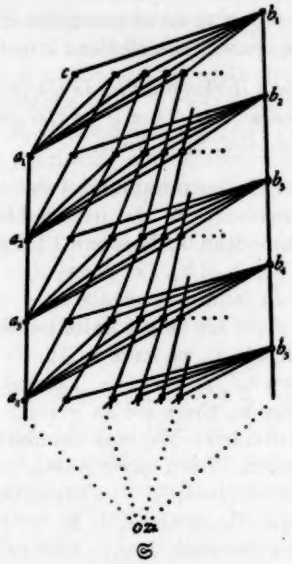


FIG. 2

$x\cap a$ is covered by an infinite number of ideals. The proof is thus complete.

Lemmas 6.2 and 6.3 and Theorem 6.2 give immediately

THEOREM 6.7. *Let each element of $\mathfrak{S}$ be covered by at most a finite number of ideals. Then the following conditions are equivalent*:

(1) E_6 *holds in* $\mathfrak{S}$.

(2) $\mathfrak{S}$ *is a Birkhoff lattice.*

(3) B1 *holds in the lattice of ideals.*

If $\mathfrak{S}$ is a Birkhoff lattice in which each element is *not* covered by at most a finite number of ideals, then even though the ascending chain condition holds in $\mathfrak{S}$ B1 need not be satisfied in the lattice of ideals. For example, consider the lattice diagramed in Figure 2.

All of the elements distinct from z form an ideal $\mathfrak{a}$ which is generated by $a_1, a_2, a_3, \cdots . b_1, b_2, b_3, \cdots$ clearly form an ideal $\mathfrak{b}$ which divides $\mathfrak{a}$. Now let $\mathfrak{b}\supset\mathfrak{c}\supset\mathfrak{a}$, $\mathfrak{b}\neq\mathfrak{c}$. Let $y\in\mathfrak{c}$, $y\notin\mathfrak{b}$. Then there exists a b_i such that $b_i>x\supset y$, $x\notin\mathfrak{b}$. But by the method of construction there exists an integer j such that

$x\cap b_k = a_k$ all $k \geqq j$. Hence $\mathfrak{a} \supset x\cap\mathfrak{b} \supset y\cap\mathfrak{b} \supset \mathfrak{c}\cap\mathfrak{b} = \mathfrak{c}$. Thus $\mathfrak{c} = \mathfrak{a}$ and $\mathfrak{b} > \mathfrak{a}$. Clearly $\mathfrak{b}\cap a_1 = \mathfrak{a}$ and $\mathfrak{b}\cup a_1 = b_1$. But then $\mathfrak{b} > a_1\cap\mathfrak{b}$ and $a_1\cup\mathfrak{b} \not> a_1$. Hence B1 does *not* hold in $\mathfrak{L}$. On the other hand it is readily verified that $\mathfrak{S}$ is a Birkhoff lattice since $\mathfrak{a}$ is the only non-principal ideal which covers a principal ideal and every element distinct from z divides $\mathfrak{a}$.

The number of ideals covering an element of a lattice is closely related to the number of decompositions of the element into irreducibles. We prove

THEOREM 6.8. *Let $\mathfrak{S}$ be a Birkhoff lattice in which each element can be represented as a cross-cut of irreducibles. Then if an element a has a finite number of decompositions into irreducibles, $\mathfrak{L}_a$ is finite.*

Proof. Since a has only a finite number of decompositions into irreducibles, the number of components in the irreducible decompositions of a is bounded. Hence $\mathfrak{L}_a$ is archimedean by Theorem 3.1. Let $\mathfrak{p}_1, \cdots, \mathfrak{p}_k$ be a maximal independent set of point ideals of $\mathfrak{L}_a$. Let $\mathfrak{s}_i = \mathfrak{p}_1\cup \cdots \cup\mathfrak{p}_{i-1}\cup\mathfrak{p}_{i+1}\cup \cdots \cup\mathfrak{p}_k$ and suppose that $\mathfrak{L}_a$ has an infinite sequence $\mathfrak{s}_1, \cdots, \mathfrak{s}_k, \mathfrak{s}_{k+1}, \cdots$ of simple ideals. Now if for each i there are only a finite number of simple ideals of the sequence which do not divide $\mathfrak{p}_i$, we have $\mathfrak{s}_n \supset \mathfrak{p}_i$ for all $n \geqq l_i$ for some l_i. Let $n \geqq \max(l_1, \cdots, l_k)$. Then $\mathfrak{s}_n \supset \mathfrak{p}_i$, $i = 1, \cdots, k$, and $\mathfrak{s}_n \supset \mathfrak{u}_a$, which is impossible. Hence for some $\mathfrak{p}_i$, say $\mathfrak{p}_k$, there are an infinity of ideals in the sequence $\mathfrak{s}_1, \mathfrak{s}_2, \cdots$ which do not divide $\mathfrak{p}_k$. We may assume that $\mathfrak{s}_k, \mathfrak{s}_{k+1}, \cdots$ do not divide $\mathfrak{p}_k$. Let $q_i \supset \mathfrak{s}_i$, $q_i \not\supset \mathfrak{u}_a$. Then $q_i \neq q_j$, $i \neq j$, since otherwise $q_i = q_i \cup q_j \supset \mathfrak{s}_i \cup \mathfrak{s}_j \supset \mathfrak{u}_a$. Now $a = \mathfrak{p}_k \cap \mathfrak{s}_{k+l} = \mathfrak{s}_1 \cap \cdots \cap \mathfrak{s}_{k-1} \cap \mathfrak{s}_{k+l}$ $(l = 0, 1, 2, \cdots)$ implies $a = q_1 \cap q_2 \cap \cdots \cap q_{k-1} \cap q_{k+l}$ $(l = 0, 1, 2, \cdots)$ by Theorem 3.3. If this representation is reduced for each l, q_{k+l} always remains since otherwise $a = q_1 \cap \cdots \cap q_{k-1} \supset \mathfrak{p}_k$. Hence a has an infinite number of irreducible decompositions, which contradicts our hypothesis. Thus $\mathfrak{L}_a$ has only a finite number of simple ideals. But by Theorem 3.2 every ideal of $\mathfrak{L}_a$ can be expressed as a cross-cut of simple ideals. Hence $\mathfrak{L}_a$ is finite.

Theorems 6.7 and 6.8 give

THEOREM 6.9. *Let $\mathfrak{S}$ be a lattice in which every element has at least one and at most a finite number of decompositions into irreducibles. Then $\mathfrak{S}$ is a Birkhoff lattice if and only if* B1 *is satisfied in the lattice of ideals*[11].

7. **Example of a Birkhoff lattice.** In §3 we have shown that the existence of a decomposition into irreducibles for an element a of a modular lattice implies that $\mathfrak{L}_a$ is archimedean. Hence if the ascending chain condition holds,

(11) Various considerations suggest that the finiteness of the number of irreducible decompositions of an element always implies the finiteness of the number of ideals covering the element, in which case E_4 and the finiteness of the number of decompositions would imply that $\mathfrak{S}$ is a Birkhoff lattice. However, I have been unable to prove this. I have also been unable to prove that E_4 is equivalent to the Birkhoff condition under the assumption of the ascending chain condition although this seems quite likely.

$\mathfrak{L}_a$ is archimedean for each a. Also if the ascending chain condition holds in a Birkhoff lattice and the number of components is bounded for an element a, then $\mathfrak{L}_a$ is archimedean. We shall construct in this section a Birkhoff lattice satisfying the ascending chain condition but containing an element a such that $\mathfrak{L}_a$ is not archimedean. By the above remark the number of components in the irreducible decompositions of a must be unbounded.

Latin capitals $A, B, C, \cdots$ will denote finite subsets of the set of positive integers $1, 2, 3, \cdots$. If A is such a set, let $n(A)$ denote the number of elements in A. Small Latin letters $a, b, c, \cdots$ will denote positive integers. $A\cup B$ and $A\cap B$ will denote set-theoretic union and cross-cut respectively and $a\cup b$, $a\cap b$ are respectively the maximum and minimum of a and b. Let $\mathfrak{S}$ be the set of all ordered couples $\alpha=\{A, a\}$ where $n(A)<a$ together with the elements u and z. In $\mathfrak{S}$ we define

$$\begin{aligned} \alpha \cup \beta &= \{A \cup B, a \cup b\} && \text{if } n(A \cup B) < a \cup b, \\ &= u && \text{if } n(A \cup B) \geqq a \cup b, \\ \alpha \cap \beta &= \{A \cap B, a \cap b\} && \text{if } A \cap B \text{ is not null}, \\ &= z && \text{if } A \cap B \text{ is null}, \\ \alpha \cup z &= \alpha \cap u = \alpha, && \alpha \cup u = u, \quad \alpha \cap z = z. \end{aligned}$$

If $A\cap B$ exists, then $n(A\cap B)\leqq n(A)<a\leqq a\cap b$. Hence $\alpha\cap\beta$ is in $\mathfrak{S}$ if α and β are in $\mathfrak{S}$. Now it can be readily verified that the union and cross-cut so defined in $\mathfrak{S}$ are idempotent, commutative and associative. Consider $\alpha\cap(\alpha\cup\beta)$. If $\alpha\cup\beta=u$, then $\alpha\cap(\alpha\cup\beta)=\alpha$. If $\alpha\cup\beta\neq u$, then $\alpha\cap(\alpha\cup\beta)=\{A\cap(A\cup B), a\cap(a\cup b)\}=\{A, a\}=\alpha$. Hence $\alpha\cap(\alpha\cup\beta)=\alpha$ in all cases. Similarly $\alpha\cup(\alpha\cap\beta)=\alpha$. Hence $\mathfrak{S}$ is a lattice under the union and cross-cut operations defined above. Clearly $\alpha\supset\beta$ if and only if $A\supset B$ and $a\leqq b$.

$\mathfrak{S}$ satisfies the ascending chain condition. For let $\alpha_1\subset\alpha_2\subset\alpha_3\subset\cdots$ be an infinite ascending chain. We may assume that $\alpha_1\neq z$ so that $\alpha_1=\{A_1, a_1\}$. But then $A_1\subset A_2\subset A_3\subset\cdots$ and $a_1\geqq a_2\geqq a_3\geqq\cdots$. Now $n(A_i)<a_i\leqq a_1$. Hence the chain $A_1\subset A_2\subset\cdots$ has only a finite number of distinct sets. Clearly the chain $a_1\geqq a_2\geqq\cdots$ has only a finite number of distinct members. Thus the chain $\alpha_1\subset\alpha_2\subset\cdots$ has only a finite number of distinct members.

Now let $\mathfrak{a}$ be an ideal of $\mathfrak{S}$ with elements $\beta, \gamma, \delta, \cdots$. Then the set of integers is either bounded or unbounded. If bounded, let a be the largest of them. Now suppose that $B\cap C\cap D\cap\cdots$ is null. Then there exist a finite number of them $B, C, \cdots, L$ whose cross-cut is null. But then $\mathfrak{a}$ contains z since $\mathfrak{a}$ is closed with respect to finite cross-cut. Hence $\mathfrak{a}=\mathfrak{S}$ in this case. If $B\cap C\cap D\cap\cdots$ is not null, let $A=B\cap C\cap D\cap\cdots$. Then $A=B\cap C\cap\cdots\cap L$ for a finite number of the sets and hence $\alpha'=\{A, a'\}$ and $\alpha''=\{A', a\}$ are in $\mathfrak{a}$. But then $\mathfrak{a}$ contains $\alpha=\alpha'\cap\alpha''=\{A, a\}$ and $\beta\supset\alpha$ for every $\beta\in\mathfrak{a}$. Hence if the integers of the ideal are bounded, $\mathfrak{a}$ is a principal ideal. If the integers of the ideal are unbounded, then as before either $\mathfrak{a}=\mathfrak{S}$

or there exists a set A such that $\{A, a\}$ is in $\mathfrak{a}$ for each positive integer a and $\beta \supset \{A, a\}$ for some a if $\beta \in \mathfrak{a}$. Hence the ideals of $\mathfrak{S}$ have the form $\mathfrak{a} = \{A, \infty\}$ if $\mathfrak{a}$ is not principal. ∞ denotes the ideal of all positive integers. Clearly if a and b are positive integers or ∞

$$\begin{aligned}
\mathfrak{a} \cap \mathfrak{b} &= \{A \cap B, a \cap b\} && \text{if } A \cap B \text{ is not null,}\\
&= z && \text{if } A \cap B \text{ is null,}\\
\mathfrak{a} \cup \mathfrak{b} &= \{A \cup B, a \cup b\} && \text{if } n(A \cup B) < a \cup b,\\
&= u && \text{if } n(A \cup B) \geqq a \cup b.
\end{aligned}$$

THEOREM 7.1. *$\mathfrak{S}$ is a Birkhoff lattice.*

Proof. We shall show that B1 holds in the lattice of ideals. Let $\mathfrak{a} > \mathfrak{a} \cap \mathfrak{b}$. If $\mathfrak{a} \cap \mathfrak{b} = z$, then $A \cap B$ is null. Since $\mathfrak{a} > z$ we have $\mathfrak{a} = \{(a), \infty\}$. But then $\mathfrak{a} \cup \mathfrak{b} = \{(a) \cup B, \infty \cup b\} = \{(a) \cup B, b\}$ and $(a) \cup B > B$. Thus $\mathfrak{a} \cup \mathfrak{b} > \mathfrak{b}$. If $\mathfrak{a} \cap \mathfrak{b} \neq z$, then $\{A, a\} > \{A \cap B, a \cap b\}$ and hence either $A > A \cap B$, $a = a \cap b$ or $A = A \cap B$, $a = (a \cap b) - 1$. In the first case $A > A \cap B \rightarrow A \cup B > B$ and $a = a \cap b$ implies $b = a \cup b$. Hence $\mathfrak{a} \cup \mathfrak{b} = \{A \cup B, a \cup b\} = \{A \cup B, b\} > \{B, b\} = \mathfrak{b}$. If $n(A \cup B) \geqq b$, then $n(B) = b - 1$ and hence $u > B$. In the second case $A = A \cap B \rightarrow A \cup B = B$ and $a > a \cap b \rightarrow a = b - 1$. Hence $\mathfrak{a} \cup \mathfrak{b} = \{B, a\} > \mathfrak{b}$. Thus $\mathfrak{a} > \mathfrak{a} \cap \mathfrak{b}$ implies $\mathfrak{a} \cup \mathfrak{b} > \mathfrak{b}$ and B1 holds in the lattice of ideals of $\mathfrak{S}$.

The point ideals of $\mathfrak{L}_s$ are clearly the ideals $\mathfrak{p}_i = \{(i), \infty\}$. Now $\mathfrak{p}_1 \cup \mathfrak{p}_2 \cup \cdots = u$ and hence $\bar{\mathfrak{L}}_s = \mathfrak{L}$. The ascending chain $\mathfrak{p}_1 \subset \mathfrak{p}_1 \cup \mathfrak{p}_2 \subset \mathfrak{p}_1 \cup \mathfrak{p}_2 \cup \mathfrak{p}_3 \subset \cdots$ has distinct members and $\mathfrak{L}_s$ is thus *not* archimedean. The point ideals $\mathfrak{p}_1, \mathfrak{p}_2, \cdots$ are *not* independent since the union of any infinite set is u. However, every finite set of the $\mathfrak{p}_i$ is independent and thus generates a Boolean algebra. $\mathfrak{L}_s$ is *not* complemented. For if $\mathfrak{p}_i \cup \mathfrak{a} = u$, then $\mathfrak{a} = u$ and $\mathfrak{a} \supset \mathfrak{p}_i$. The simple ideals of $\mathfrak{L}_s$ are the elements of the form $\{A, a\}$ where $n(A) = a - 1$. Clearly $\mathfrak{p}_i$ cannot be represented as a cross-cut of simple ideals. Hence it is *not* true that every ideal of $\mathfrak{L}_s$ may be represented as a cross-cut of simple ideals. The irreducibles of $\mathfrak{S}$ are the simple elements of $\mathfrak{L}_s$, namely, those elements $\{A, a\}$ with $n(A) = a - 1$. Now let A_i be the set $\{1, 2, \cdots, i-1, i+1, \cdots, k\}$ $(i = 1, \cdots, k)$. Then $\alpha_i = \{A_i, k\}$ is simple for each i and $z = \alpha_1 \cap \alpha_2 \cap \cdots \cap \alpha_k$ since $A_1 \cap A_2 \cap \cdots \cap A_k$ is null. This representation of z is clearly reduced. Hence for any positive integer $k > 1$, z has a reduced decomposition with k components.

This example clearly indicates the complications that may arise if $\mathfrak{L}_a$ is *not* archimedean even though the ascending chain condition holds in $\mathfrak{S}$.

8. **Example of a lattice satisfying E_5 which is not a Birkhoff lattice.** Let S be the set of elements $p_1, p_2, p_3, \cdots$. From the set of all subsets of S omit those infinite sets which contain either p_1 or p_2 but not both. Denote this set of subsets by $\mathfrak{S}$. $\mathfrak{S}$ is clearly closed under infinite cross-cut. Since $\mathfrak{S}$ contains a unit element, the union of any set of sets of $\mathfrak{S}$ may be defined in terms

of the cross-cut operation. $\mathfrak{S}$ is thus a continuous lattice in which every element is a union of points. Now consider $A\cup p$ where $A\in\mathfrak{S}$ and p is any element of S. If A is finite, then clearly $A\cup p=A+p$ where $+$ indicates set-theoretic union. Also if A is infinite and contains both p_1 and p_2, $A\cup p=A+p$. Now if A is infinite and does not contain p_1 or p_2, then $A\cup p=A+p$ if $p\neq p_1, p_2$ and $A\cup p=A+p_1+p_2$ if $p=p_1$ or $p=p_2$. Now let $A+p_1+p_2\supset B\supset A$. If $B\neq A$, then B contains p_1 or p_2 and hence contains both p_1 and p_2 by the definition of $\mathfrak{S}$. Thus $B=A+p_1+p_2$. Hence in every case $A\cup p>A$ if $p\notin A$.

Now let $A\supset B\supset A\cap C$ and $C\neq A\cap C$ where A, B, C are in $\mathfrak{S}$. Since every element of $\mathfrak{S}$ is a union of points and $C\neq A\cap C$, there exists a point p such that $C\supset p$, $A\cap C\not\supset p$. Set $C_1=(A\cap C)\cup p$. Then $C\supset C_1\supset A\cap C$ and $C_1\neq A\cap C$. Now $B\cup p\supset A\cap(B\cup C_1)\supset B$ and $B\cup p\neq A\cap(B\cup C_1)$ since otherwise $A\supset B\cup p\supset p$ and $A\cap C\supset p$ which contradicts $A\cap C\not\supset p$. Since $B\cup p>B$ we thus have $B=A\cap(B\cup C_1)$ and hence E_5 holds in $\mathfrak{S}$.

In $\mathfrak{S}$ let $\mathfrak{a}$ be the ideal generated by the sets $A_k=\{p_k, p_{k+1}, \cdots\}$ $(k=3, 4, 5, \cdots)$. Then by Theorem 2.1, there exists a point ideal $\mathfrak{p}$ such that $\mathfrak{a}\supset\mathfrak{p}>\mathfrak{z}$. Every set of $\mathfrak{S}$ occurring in $\mathfrak{p}$ contains an infinite number of elements. For suppose that $Q\in\mathfrak{p}$ and Q contains only a finite number of elements. Let k be the largest subscript occurring among the elements of Q. Then $\mathfrak{z}=Q\cap A_{k+1}\supset\mathfrak{p}\cap\mathfrak{a}\supset\mathfrak{p}$ which contradicts $\mathfrak{p}>\mathfrak{z}$. If $Q\in\mathfrak{p}$, then $Q\cup p_1\supset p_2$ by the definition of $\mathfrak{S}$. Hence $\mathfrak{p}\cup p_1\supset p_1\cup p_2\supset p_1$ where $\mathfrak{p}\cup p_1\neq p_1\cup p_2$ and $p_1\cup p_2\neq p_1$. Thus $\mathfrak{p}\cup p_1\not> p_1$ and hence $\mathfrak{S}$ is not a Birkhoff lattice. If $A\supset B$ and $A\neq B$, let $A\supset p$, $B\not\supset p$. Then $A\supset B_1>B$ where $B_1=B\cup p$. Thus $\mathfrak{S}$ is an example of a continuous point lattice in which covering elements exist and E_5 holds, but which is *not* a Birkhoff lattice. Whether or not an exchange lattice, i.e., a continuous point lattice satisfying E_5 and a finite dependence axiom (Mac Lane [1]) is a Birkhoff lattice is an open question.

References

G. Birkhoff

1. *Rings of sets*, Duke Mathematical Journal, vol. 3 (1937), pp. 443-454.
2. *On combination of subalgebras*, Proceedings of the Cambridge Philosophical Society, vol. 29 (1933), pp. 441–464.

R. P. Dilworth

1. *Lattices with unique irreducible decompositions*, Annals of Mathematics.
2. *The arithmetical theory of Birkhoff lattices*, submitted to Duke Mathematical Journal.

S. Mac Lane

1. *A lattice formulation for transcendence degrees and p-bases*, Duke Mathematical Journal, vol. 24 (1938), pp. 455-468.

M. H. Stone

1. *The theory of representations of Boolean algebra*, these Transactions, vol. 40 (1936), pp. 37–111.

H. Wallman

1. *Lattices and topological spaces*, Annals of Mathematics, (2), vol. 39 (1938), pp. 112–126.

Yale University,
New Haven, Conn.

GENERALIZATIONS TO SPACE OF THE CAUCHY AND MORERA THEOREMS[1]

BY

MAXWELL READE AND E. F. BECKENBACH

INTRODUCTION

Consider the function

$$w = f(z) = x_1(u, v) + ix_2(u, v), \qquad z = u + iv, \tag{1}$$

defined and continuous in a simply connected domain D[2]. A necessary and sufficient condition that $f(z)$ be analytic in D is that the Cauchy-Riemann equations be satisfied there:

$$\lambda w = 0, \tag{2}$$

where

$$\lambda \equiv \frac{\partial}{\partial u} + i \frac{\partial}{\partial v}$$

is a differential operator. From (2) we obtain

$$\sum_{j=1}^{2} (\lambda x_j)^2 = 0. \tag{3}$$

According to the Cauchy and Morera theorems, a necessary and sufficient condition that the continuous function (1) be analytic in the simply connected domain D is that for each closed rectifiable Jordan curve γ lying in the domain D,

$$\int_\gamma f(z)dz = 0. \tag{4}$$

Now (4) may be considered to be an integral analogue of the differential condition (2); it implies

$$\sum_{j=1}^{2} \left[\int_\gamma x_j(u, v)dz \right]^2 = 0, \tag{5}$$

which is analogous to (3).

Presented to the Society April 14, 1939, under the title *A characterization of plane isothermic maps*; received by the editors March 4, 1940.

(1) Some of the results of this paper have been summarized in the authors' note of the same title in the Proceedings of National Academy of Sciences, vol. 25 (1939), pp. 92–97.

(2) It is to be understood throughout this paper that the domains under consideration are finite.

If (3) holds, then either x_1+ix_2 or x_2+ix_1 is an analytic function of $z=u+iv$, and $x_1(u, v)$ and $x_2(u, v)$ are said to form *a couple of conjugate harmonic functions*. We note that (5) is a necessary and sufficient condition that the continuous functions $x_1(u, v)$ and $x_2(u, v)$ form a couple of conjugate harmonic functions.

The real functions

$$x_j = x_j(u, v), \qquad j = 1, 2, 3, \tag{6}$$

defined and continuous in a simply connected domain D, will be said to define a surface S. If the first partial derivatives of the functions (6) are continuous and satisfy

$$E(u, v) = G(u, v), \qquad F(u, v) = 0, \tag{7}$$

in D, where

$$E(u, v) \equiv \sum_{j=1}^{3}\left(\frac{\partial x_j}{\partial u}\right)^2, \quad F(u, v) \equiv \sum_{j=1}^{3}\frac{\partial x_j}{\partial u}\frac{\partial x_j}{\partial v}, \quad G(u, v) \equiv \sum_{j=1}^{3}\left(\frac{\partial x_j}{\partial v}\right)^2 \tag{8}$$

are the coefficients of the first fundamental differential quadratic form of the surface S, then the parameters u, v are said to be isothermic parameters, and S is said to be given in isothermic representation. The map of D on S is conformal except where $E=G=0$. From (8) it follows that (7), which is a generalization to space of (3), can be written in the form

$$\sum_{j=1}^{3}(\lambda x_j)^2 = 0. \tag{9}$$

An analogous generalization to space of (5) is

$$\sum_{j=1}^{3}\left[\int_\gamma x_j(u, v)dz\right]^2 = 0, \tag{10}$$

where γ is a closed rectifiable Jordan curve lying in D. In this paper we shall study (10) and

$$\sum_{j=1}^{3}\left[\int_{C_r} x_j(u, v)dz\right]^2 = o(r^\alpha),$$

where C_r is the circle in D with center at the arbitrary point (u_0, v_0) of D and of radius r, and where $o(r^\alpha)$ denotes a quantity (not always the same quantity) such that

$$\lim_{r\to 0}\frac{o(r^\alpha)}{r^\alpha} = 0.$$

If the functions (6) are harmonic and satisfy (9) in a simply connected

domain D, they have been called *a triple of conjugate harmonic functions*([3]). In terms of this definition, a theorem of Weierstrass may be stated as follows.

A necessary and sufficient condition that the functions (6), *defined in a simply connected domain, be the coordinate functions of a minimal surface given in isothermic representation is that they form a triple of conjugate harmonic functions.*

We shall have use for the following direct computation. Let the functions (6) have continuous partial derivatives of the mth order in a simply connected domain D; then about each point (u_0, v_0) of D we have a finite Taylor expansion for each function:

$$x_j(u, v) = \sum_{n=0}^{m} \frac{r^n}{n!}\left[\left(\cos\theta \frac{\partial}{\partial u} + \sin\theta \frac{\partial}{\partial v}\right)^n x_j\right] + o(r^m), \qquad j = 1, 2, 3, \tag{11}$$

where

$$\cos\theta \frac{\partial}{\partial u} + \sin\theta \frac{\partial}{\partial v}$$

is a differential operator, where the partial derivatives are evaluated at the point (u_0, v_0), and where $u-u_0=r\cos\theta$, $v-v_0=r\sin\theta$. *Then, for the circle C_r in D with center at (u_0, v_0) and radius r, the left-hand member of* (10) *assumes the form*

$$\sum_{j=1}^{3}\left[\int_{C_r} x_j(u, v)dz\right]^2 = -\pi^2 \sum_{k=0}^{[(m-1)/2]} \sum_{p=0}^{k} r^{2k+4}B_{p,k} + o(r^{m+3}), \tag{12}$$

where

$$B_{p,k} \equiv \sum_{j=1}^{3} \frac{C_{k+1,p+1}C_{k+1,p}}{2^{2k}[(k+1)!]^2}\Delta^p\lambda x_j\Delta^{k-p}\lambda x_j,$$

where

$$\Delta \equiv \lambda\bar{\lambda} \equiv \frac{\partial^2}{\partial u^2} + \frac{\partial^2}{\partial v^2}, \qquad \Delta^0\lambda x_j \equiv \lambda x_j,$$

where the $C_{k,s}$ are binomial coefficients, $C_{k,s}\equiv k!/s!(k-s)!$, $C_{k,0}\equiv 1$, and where $[(m-1)/2]$ is the greatest integer not greater than $(m-1)/2$.

For $m=1, 3, 5$, (12) is displayed in (14), (48) and (35) respectively.

1. Characterization of isothermic maps

THEOREM 1. *If the functions* (6) *have continuous partial derivatives of the first order in a simply connected domain D, then a necessary and sufficient condition that they map D isothermically on a surface S is that for each point (u_0, v_0) of D,*

([3]) E. F. Beckenbach and T. Radó, *Subharmonic functions and minimal surfaces*, these Transactions, vol. 35 (1933), pp. 648–661.

$$\sum_{j=1}^{3}\left[\int_{C_r} x_j(u, v)dz\right]^2 = o(r^4), \tag{13}$$

where C_r is the circle in D with center at (u_0, v_0) and radius r.

Proof. If the first partial derivatives of the functions (6) are continuous in D, then we obtain a finite Taylor expansion for each function about an arbitrary point (u_0, v_0) of D by setting $m=1$ in (11). If C_r is the circle in D with center at (u_0, v_0) and radius r, then upon setting $m=1$ in (12) we obtain

$$\sum_{j=1}^{3}\left[\int_{C_r} x_j(u, v)dz\right]^2 = -\pi^2 r^4 \sum_{j=1}^{3} (\lambda x_j)^2 + o(r^4). \tag{14}$$

From (14) it follows that a necessary and sufficient condition that the relation (9) hold is that (13) hold.

2. Characterization of those isothermic spherical maps that do not map circles on circles

LEMMA 1. *If the functions* (6) *are defined, but not identically constant, in a simply connected domain D, and if they map D isothermically on a surface S that lies on a sphere $\mathfrak{S}$ of finite radius r, then a necessary and sufficient condition that they map circles on circles is that the first quadratic form of S have the representation*

$$ds^2 = \frac{c^2(du^2 + dv^2)}{[(u-u_0)^2 + (v-v_0)^2 + a^2]^2}, \qquad a > 0,\ c > 0. \tag{15}$$

Necessity. Let the coordinates of the center of $\mathfrak{S}$ be denoted by (a_1, a_2, a_3), and let an s, t-plane be so placed that the positive s- and t-axes coincide with the positive x_1- and x_2-axes respectively. Let $\mathfrak{S}$ be projected stereographically on the s, t-plane, the coordinates of the pole of projection being (a_1, a_2, k), where $k=a_3 \pm r$ and $|k|$ is the maximum of the two quantities $|a_3+r|$ and $|a_3-r|$. We have

$$\begin{aligned} x_1 &= a_1 + \frac{2r|k|(s-a_1)}{(s-a_1)^2 + (t-a_2)^2 + k^2}, \\ x_2 &= a_2 + \frac{2r|k|(t-a_2)}{(s-a_1)^2 + (t-a_2)^2 + k^2}, \\ x_3 &= k\left[1 - \frac{2r|k|}{(s-a_1)^2 + (t-a_2)^2 + k^2}\right]. \end{aligned} \tag{16}$$

Let the map of S on the s, t-plane be D'. The product of the transformation (6) and the stereographic projection maps D on D', and carries circles in D into circles in D'. It follows that this isothermic map of D on D' is equivalent

to a single linear transformation:

$$F(z) = s + it = \frac{\alpha' z + \beta'}{\gamma' z + \delta'}, \tag{17a}$$

or

$$F(z) = s + it = \frac{\alpha' \bar{z} + \beta'}{\gamma' \bar{z} + \delta'}, \qquad (\alpha'\delta' - \beta'\gamma') \neq 0, \tag{17b}$$

where (a) holds if the map of D on D' is directly conformal, and where (b) holds if the map is inversely conformal. From (16) and (17) it follows that the functions (6) have the form

$$\begin{gathered} x_1 = a_1 + \frac{2r|k|\,\mathfrak{R}f(z)}{f(z)\overline{f(z)} + k^2}, \qquad x_2 = a_2 + \frac{2r|k|\,\mathfrak{I}f(z)}{f(z)\overline{f(z)} + k^2}, \\ x_3 = k\left[1 - \frac{2r|k|}{f(z)\overline{f(z)} + k^2}\right], \end{gathered} \tag{18}$$

where

$$f(z) = F(z) - (a_1 + ia_2), \tag{19}$$

and where $\mathfrak{R}f(z)$ is the real part, and $\mathfrak{I}f(z)$ the imaginary part, of $f(z)$.

There are two possible representations for the functions (6) as determined by (17), (18) and (19). It is now an easy matter to compute E, F and G and to show that the first quadratic form of S has the representation (15)([4]).

We note that *if it is given only that one non-null circle in D is mapped on a circle on S by the isothermic functions* (6), then, as in the above discussion, D is mapped isothermically on D' such that one non-null circle in D is mapped on a circle in D'. Therefore the function mapping D isothermically on D' is linear; and hence *it follows that the functions* (6) *have the form* (18) *and map all circles in D on circles on S.*

Sufficiency. Let the stereographic projection of D on the sphere $\mathfrak{S}'$ be S', where $\mathfrak{S}'$ is the sphere with center at $(u_0, v_0, a - c/2a)$ and radius $c/2a$, and where the pole of projection is at (u_0, v_0, a). This projection is given by (18), where

$$a_1 = u_0, \qquad a_2 = v_0, \qquad k = a, \qquad r = c/2a, \qquad f(z) = (u + iv) - (u_0 + iv_0).$$

The first quadratic form of S' is found to be identical with that of S; hence S' and S are applicable. Therefore S' and S are either congruent or symmetric;

([4]) We find $ds^2 = E(du^2 + dv^2) = c^2(du^2 + dv^2)/[(u - u_0)^2 + (v - v_0)^2 + \eta]^2$. But then a computation shows that the Gaussian curvature, given by the formula in (47), which is positive on $\mathfrak{S}$, has the value $4\eta/c^2$, so that $\eta = a^2$, where $a > 0$.

in either case since circles in D are mapped on circles on S', it follows that circles in D are mapped on circles on S.

THEOREM 2. *If the functions* (6) *are defined, but not identically constant, in a simply connected domain D, and if they have continuous partial derivatives of the third order in D, then a necessary and sufficient condition that they map D isothermically on a surface S that lies on a sphere of finite radius, such that circles are not mapped on circles, is that*

$$\sum_{j=1}^{3}\left[\int_{C_r} x_j(u, v)dz\right]^2 = o(r^6) \tag{20}$$

hold for all points (u_0, v_0) in D, while

$$\sum_{j=1}^{3}\left[\int_{C_r} x_j(u, v)dz\right]^2 \neq o(r^8) \tag{21}$$

hold for some points (u_0, v_0) in D, where C_r is the circle in D with center at (u_0, v_0) and radius r.

Necessity. Under the hypotheses, the functions (6) map D isothermically on a surface that lies on a sphere of finite non-null radius. Hence (9) holds, and there exists a constant α such that

$$e = \alpha E, \qquad f = \alpha F, \qquad g = \alpha G, \qquad \alpha \neq 0, \tag{22}$$

where e, f and g are the coefficients of the second fundamental quadratic form of S(5).

For the present representation of S, the formulas of Gauss(6) become

$$\begin{aligned} x_{j,uu} &= R_u x_{j,u} - R_v x_{j,v} + e\zeta_j, \qquad x_{j,uv} = R_v x_{j,u} + R_u x_{j,v},\\ x_{j,vv} &= - R_u x_{j,u} + R_v x_{j,v} + e\zeta_j, \qquad j = 1, 2, 3, \end{aligned} \tag{23}$$

where

$$R \equiv \tfrac{1}{2} \log E,$$

and where ζ_j, $j=1, 2, 3$, are the direction cosines of the normal to S.

We shall need the following relations:

$$\sum_{j=1}^{3} \lambda x_j \Delta^2 \lambda x_j = - 4\alpha^2 E\lambda^2 E, \tag{24}$$

$$\sum_{j=1}^{3} (\Delta\lambda x_j)^2 = 4\alpha^2(\lambda E)^2. \tag{25}$$

(5) For information concerning the second fundamental differential quadratic form of a surface, see W. C. Graustein, *Differential Geometry*; in particular, we have referred to pages 93–94 and pages 97–98.

(6) Graustein, op. cit., pp. 135–137.

To obtain (24) and (25), we shall use a method which depends upon the existence and continuity of partial derivatives of order higher than three of the functions (6). These functions map D isothermically on a spherical surface. Therefore, as in the proof of Lemma 1, we can consider an intermediate stereographic projection to show that the functions (6) have the representation (18), where either $f(z)$ or $f(\bar{z})$ is analytic in D. From this representation, it follows that the functions (6) have continuous partial derivatives of all orders.

To obtain (24) and (25) we shall need the following equalities:

$$\Delta\lambda x_j = 2(e\lambda\zeta_j + \zeta_j\lambda e), \qquad j = 1, 2, 3, \tag{26}$$

$$\Delta^2\lambda x_j = 2(e\Delta\lambda\zeta_j + 2\lambda e\Delta\zeta_j + 2\Delta e\lambda\zeta_j + \bar{\lambda}e\lambda^2\zeta_j + \lambda^2 e\bar{\lambda}\zeta_j + \zeta_j\Delta\lambda e), \qquad j = 1, 2, 3, \tag{27}$$

$$\sum_{j=1}^{3} \zeta_j\lambda x_j = 0, \tag{28}$$

$$\sum_{j=1}^{3} \lambda x_j\lambda\zeta_j = 0, \tag{29}$$

$$\sum_{j=1}^{3} \lambda^2\zeta_j\lambda x_j = 0, \tag{30}$$

$$\sum_{j=1}^{3} \lambda x_j\Delta\zeta_j = 0, \tag{31}$$

$$\sum_{j=1}^{3} \lambda x_j\Delta\lambda\zeta_j = 0. \tag{32}$$

It follows from (23) that

$$\Delta\lambda x_j = 2\lambda(e\zeta_j), \qquad j = 1, 2, 3,$$

and hence (26) holds. Operating on (26) with the operator $\bar{\lambda}$, and then operating on this result with λ, we obtain (27).

The formulas of Olinde Rodriguez(7) may be written together in the following form :

$$\lambda\zeta_j = -\alpha\lambda x_j, \qquad \alpha \neq 0, j = 1, 2, 3. \tag{33}$$

From (33) we obtain

$$\sum_{j=1}^{3} \lambda x_j\lambda\zeta_j = -\alpha\sum_{j=1}^{3} (\lambda x_j)^2,$$

which, with (9), yields (29).

From (29) we obtain

(7) Graustein, op. cit., p. 121.

$$\lambda\sum_{j=1}^{3}\lambda\zeta_j\lambda x_j = \sum_{j=1}^{3}\lambda^2\zeta_j\lambda x_j + \sum_{j=1}^{3}\lambda\zeta_j\lambda^2 x_j = 0,$$

which, with (33), establishes (30).

From (29) we obtain

$$\bar{\lambda}\sum_{j=1}^{3}\lambda\zeta_j\lambda x_j = \sum_{j=1}^{3}\Delta\zeta_j\lambda x_j + \sum_{j=1}^{3}\lambda\zeta_j\Delta x_j = 0,$$

which, with (33), establishes (31).

From (26), (28) and (29) we obtain

$$\sum_{j=1}^{3}\lambda x_j\Delta\lambda x_j = 0,$$

which, with (33), yields (32).

From (27)–(32), inclusive, it follows that

$$\sum_{j=1}^{3}\lambda x_j\Delta^2\lambda x_j = 2\lambda^2 e\sum_{j=1}^{3}\lambda x_j\bar{\lambda}\zeta_j. \tag{34}$$

From (8), (9), (22), (33) and (34) we obtain (24).

From (9) and (33) we obtain

$$\sum_{j=1}^{3}(\lambda\zeta_j)^2 = 0,$$

which combines with (22), (26) and (33) to yield (25).

Since the functions (6) have continuous partial derivatives of all orders in D, the following expression, obtained from (12) by setting $m=5$, is valid:

$$\begin{aligned}\sum_{j=1}^{3}\left[\int_{C_r} x_j(u, v)dz\right]^2 = &-\pi^2r^4\sum_{j=1}^{3}(\lambda x_j)^2 - \frac{\pi^2r^6}{4}\sum_{j=1}^{3}\lambda x_j\Delta\lambda x_j \\ &- \frac{\pi^2r^8}{192}\sum_{j=1}^{3}\left[2\lambda x_j\Delta^2\lambda x_j + 3(\Delta\lambda x_j)^2\right] + o(r^8).\end{aligned} \tag{35}$$

Here the partial derivatives are evaluated at the point (u_0, v_0) which is the center of the arbitrary circle C_r in D. Applying (9) and the above relation $\sum_{j=1}^{3}\lambda x_j\Delta\lambda x_j=0$, to (35), we obtain

$$\sum_{j=1}^{3}\left[\int_{C_r} x_j(u, v)dz\right]^2 = -\frac{\pi^2r^8}{192}\sum_{j=1}^{3}\left[2\lambda x_j\Delta^2\lambda x_j + 3(\Delta\lambda x_j)^2\right] + o(r^8). \tag{36}$$

Therefore (20) holds.

Let us suppose that (21) does not hold, i.e., that for each point (u_0, v_0) in D

$$(37)\qquad \sum_{j=1}^{3}\left[\int_{C_r} x_j(u, v)dz\right]^2 = o(r^8),$$

where C_r is the circle in D with center at (u_0, v_0) and radius r. From (36) and (37) it follows that

$$(38)\qquad \sum_{j=1}^{3} [2\lambda x_j\Delta^2\lambda x_j + 3(\Delta\lambda x_j)^2] = 0.$$

From (24), (25) and (38) we obtain

$$(39)\qquad \alpha^2[2E\lambda^2E - 3(\lambda E)^2] = 0.$$

But, by (22), $\alpha\neq 0$, and therefore (39) yields

$$(40)\qquad 2E\lambda^2E - 3(\lambda E)^2 = 0.$$

From (40) we obtain

$$2\lambda \log (\lambda E) = 3\lambda(\log E);$$

therefore

$$(41)\qquad \lambda E = E^{3/2}e^{Q(z)},$$

where

$$(42)\qquad \lambda Q(z) = 0.$$

From the imaginary part of (40) we obtain

$$2\frac{E_{uv}}{E_u}dv = 3\frac{E_v}{E}dv, \qquad 2\frac{E_{uv}}{E_v}du = 3\frac{E_u}{E}du,$$

which imply

$$(43)\qquad E_u = E^{3/2}e^{\Phi_1(u)} \qquad E_v = E^{3/2}e^{\Phi_2(v)}$$

where $\Phi_1(u)$ is a function of u alone and $\Phi_2(v)$ is a function of v alone. From (41), (42) and (43) it follows that the function

$$(44)\qquad e^{Q(z)} \equiv e^{\Phi_1(u)} + ie^{\Phi_2(v)}$$

is an analytic function of $z=u+iv$. From the Cauchy-Riemann equations for the function (44) it follows that

$$e^{\Phi_1(u)} \equiv 2a_0u + a_1, \qquad e^{\Phi_2(v)} \equiv 2a_0v + a_2,$$

where a_0, a_1 and a_2 are real constants, and hence that (41) yields

$$(45)\qquad \frac{E_u}{E^{3/2}}du = (2a_0u + a_1)du, \qquad \frac{E_v}{E^{3/2}}dv = (2a_0v + a_2)dv.$$

From (45) we obtain

$$E = \frac{4}{[a_0(u^2 + v^2) + a_1 u + a_2 v + a_3]^2}, \tag{46}$$

where a_3 is another real constant.

The Gaussian curvature of S is given by

$$K = -\frac{1}{2E}\Delta \log E(^8). \tag{47}$$

From (46) and (47) it follows that

$$K = \frac{4a_0a_3 - a_1^2 - a_2^2}{4},$$

which implies that $a_0 \neq 0$, since S is on a sphere of finite non-null radius. Since (9) holds, it follows from (7) and (46) that the first quadratic form of S may be written as follows:

$$ds^2 = \frac{\alpha^2(du^2 + dv^2)}{\left[\left(u + \frac{a_1}{2a_0}\right)^2 + \left(v + \frac{a_2}{2a_0}\right)^2 + \beta^2\right]^2},$$

where

$$\alpha = |\, 2/a_0 \,|, \qquad \beta = (K/a_0^2)^{1/2}.$$

Therefore the first quadratic form of S has the representation (15); hence it follows from Lemma 1 that all circles in D are mapped on circles on S by the functions (6). This is a contradiction of the hypothesis. Hence (21) holds for some points (u_0, v_0) in D.

Sufficiency. By setting $m=3$ in (12), we obtain

$$\sum_{j=1}^{3}\left[\int_{C_r} x_j(u, v)dz\right]^2 = -\pi^2 r^4 \sum_{j=1}^{3}(\lambda x_j)^2 - \frac{\pi^2 r^6}{4}\sum_{j=1}^{3}\lambda x_j \Delta\lambda x_j + o(r^6), \tag{48}$$

where C_r is the circle in D with center at (u_0, v_0) and radius r, and where the partial derivatives are evaluated at (u_0, v_0). From (20) and (48) it follows that (9) and

$$\sum_{j=1}^{3}\lambda x_j \Delta\lambda x_j = 0 \tag{49}$$

hold. Operating on (9) with the operator $\bar{\lambda}$ we obtain

(8) E. F. Beckenbach and T. Radó, *Subharmonic functions and surfaces of negative curvature*, these Transactions, vol. 35 (1933), pp. 662–674.

(37) $$\sum_{j=1}^{3}\left[\int_{C_r} x_j(u, v)dz\right]^2 = o(r^8),$$

where C_r is the circle in D with center at (u_0, v_0) and radius r. From (36) and (37) it follows that

(38) $$\sum_{j=1}^{3} [2\lambda x_j\Delta^2\lambda x_j + 3(\Delta\lambda x_j)^2] = 0.$$

From (24), (25) and (38) we obtain

(39) $$\alpha^2[2E\lambda^2E - 3(\lambda E)^2] = 0.$$

But, by (22), $\alpha\neq 0$, and therefore (39) yields

(40) $$2E\lambda^2E - 3(\lambda E)^2 = 0.$$

From (40) we obtain

$$2\lambda \log (\lambda E) = 3\lambda(\log E);$$

therefore

(41) $$\lambda E = E^{3/2}e^{Q(z)},$$

where

(42) $$\lambda Q(z) = 0.$$

From the imaginary part of (40) we obtain

$$2\frac{E_{uv}}{E_u}dv = 3\frac{E_v}{E}dv, \qquad 2\frac{E_{uv}}{E_v}du = 3\frac{E_u}{E}du,$$

which imply

(43) $$E_u = E^{3/2}e^{\Phi_1(u)} \qquad E_v = E^{3/2}e^{\Phi_2(v)}$$

where $\Phi_1(u)$ is a function of u alone and $\Phi_2(v)$ is a function of v alone. From (41), (42) and (43) it follows that the function

(44) $$e^{Q(z)} \equiv e^{\Phi_1(u)} + ie^{\Phi_2(v)}$$

is an analytic function of $z=u+iv$. From the Cauchy-Riemann equations for the function (44) it follows that

$$e^{\Phi_1(u)} \equiv 2a_0u + a_1, \qquad e^{\Phi_2(v)} \equiv 2a_0v + a_2,$$

where a_0, a_1 and a_2 are real constants, and hence that (41) yields

(45) $$\frac{E_u}{E^{3/2}}du = (2a_0u + a_1)du, \qquad \frac{E_v}{E^{3/2}}dv = (2a_0v + a_2)dv.$$

From (45) we obtain

$$(46)\qquad E = \frac{4}{[a_0(u^2+v^2) + a_1u + a_2v + a_3]^2},$$

where a_3 is another real constant.

The Gaussian curvature of S is given by

$$(47)\qquad K = -\frac{1}{2E}\Delta \log E(^8).$$

From (46) and (47) it follows that

$$K = \frac{4a_0a_3 - a_1^2 - a_2^2}{4},$$

which implies that $a_0 \neq 0$, since S is on a sphere of finite non-null radius. Since (9) holds, it follows from (7) and (46) that the first quadratic form of S may be written as follows:

$$ds^2 = \frac{\alpha^2(du^2+dv^2)}{\left[\left(u+\frac{a_1}{2a_0}\right)^2 + \left(v + \frac{a_2}{2a_0}\right)^2 + \beta^2\right]^2},$$

where

$$\alpha = |2/a_0|, \qquad \beta = (K/a_0^2)^{1/2}.$$

Therefore the first quadratic form of S has the representation (15); hence it follows from Lemma 1 that all circles in D are mapped on circles on S by the functions (6). This is a contradiction of the hypothesis. Hence (21) holds for some points (u_0, v_0) in D.

Sufficiency. By setting $m=3$ in (12), we obtain

$$(48)\qquad \sum_{j=1}^{3}\left[\int_{C_r} x_j(u,v)dz\right]^2 = -\pi^2r^4\sum_{j=1}^{3}(\lambda x_j)^2 - \frac{\pi^2r^6}{4}\sum_{j=1}^{3}\lambda x_j\Delta\lambda x_j + o(r^6),$$

where C_r is the circle in D with center at (u_0, v_0) and radius r, and where the partial derivatives are evaluated at (u_0, v_0). From (20) and (48) it follows that (9) and

$$(49)\qquad \sum_{j=1}^{3}\lambda x_j\Delta\lambda x_j = 0$$

hold. Operating on (9) with the operator $\bar{\lambda}$ we obtain

(8) E. F. Beckenbach and T. Radó, *Subharmonic functions and surfaces of negative curvature*, these Transactions, vol. 35 (1933), pp. 662–674.

$$\sum_{j=1}^{3} \lambda x_j \Delta x_j = 0. \tag{50}$$

Operating on (50) with λ, and applying (49) to the result, we obtain

$$\sum_{j=1}^{3} \lambda^2 x_j \Delta x_j = 0. \tag{51}$$

The four real linear homogeneous equations in Δx_j, $j=1, 2, 3$, implied by (50) and (51) are

$$\begin{gathered} \sum_{j=1}^{3} x_{j,u}\Delta x_j = 0, \qquad \sum_{j=1}^{3} x_{j,v}\Delta x_j = 0, \\ \sum_{j=1}^{3} (x_{j,uu} - x_{j,vv})\Delta x_j = 0 \qquad \sum_{j=1}^{3} (x_{j,uv})\Delta x_j = 0. \end{gathered} \tag{52}$$

One solution of (52) is

$$\Delta x_j \equiv 0, \qquad j = 1, 2, 3, \tag{53}$$

which, by (9) and the theorem of Weierstrass, implies that the functions (6) map D isothermically on a minimal surface. From (9), (12) and (53) it follows that

$$\sum_{j=1}^{3}\left[\int_C x_j(u, v)dz\right]^2 = 0 \tag{54}$$

for all circles C in D. Therefore (37) holds. This leads to a contradiction of (21); hence the functions (6) do not satisfy (53). It follows that the given functions, which satisfy (52) and do not satisfy (53), must be functions for which the rank of the matrix

$$\left\| \begin{matrix} x_{1,u} & x_{2,u} & x_{3,u} \\ x_{1,v} & x_{2,v} & x_{3,v} \\ x_{1,uv} & x_{2,uv} & x_{3,uv} \\ x_{1,uu} - x_{1,vv} & x_{2,uu} - x_{2,vv} & x_{3,uu} - x_{3,vv} \end{matrix} \right\|$$

is less than three; hence, from the definitions of e, f, g, we obtain $e=g$, $f=0$, which, with (9), imply that the functions (6) are the coordinate functions of a surface S that lies on a sphere $\mathfrak{S}$ [9] of finite[10] non-null radius.

---

(9) It is easily seen that the hypothesis of Theorem 3 on page 98, Graustein, op. cit., is satisfied.

(10) If the surface S were a plane surface, then the functions (6) would be the isothermic coordinate functions of a minimal surface. We have already considered functions (6) that satisfy (9) and (53).

Suppose that the functions (6) map D isothermically on the spherical surface S such that circles are mapped on circles[11]. Then the map of an arbitrary fixed circle C' in D is a circle, C^*, on S. Project S stereographically on the plane p of C^* and let the map of S on p be D^*. It follows that D has been mapped isothermically on the plane surface D^*. By the theorem of Weierstrass, the mapping functions form a triple of conjugate harmonic functions,

$$y_j = y_j(u, v), \qquad j = 1, 2, 3,$$

where (9) and $\Delta y_j \equiv 0$, $j=1, 2, 3$, hold. Just as for (54) it follows that

$$\sum_{j=1}^{3}\left[\int_{C'} y_j(u, v)dz\right]^2 = 0 \tag{55}$$

for the *arbitrary* fixed circle C' in D. But, for (u, v) on C',

$$x_j(u, v) = y_j(u, v), \qquad j = 1, 2, 3,$$

which, with (55) implies that (54) holds. But this leads to a contradiction of (21). Therefore the functions (6) map D isothermically on the spherical surface S such that *no* non-null circle in D is mapped on a circle on S.

3. Characterization of minimal surfaces in isothermic representation and of those isothermic spherical maps that map circles on circles

THEOREM 3. *If the functions* (6) *have continuous partial derivatives of the third order in a simply connected domain D, then a necessary and sufficient condition that they either* (1) *be the coordinate functions of a minimal surface in isothermic representation, or* (2) *map D isothermically on a surface S that lies on a sphere of finite non-null radius such that circles are mapped on circles, is that for each circle C in D*

$$\sum_{j=1}^{3}\left[\int_{C} x_j(u, v)dz\right]^2 = 0. \tag{56}$$

Necessity. We have already shown that if the functions (6) are the coordinate functions of a minimal surface in isothermic representation, then (56), i.e., (54) holds. We have also shown that if the functions (6) map D isothermically on a surface S that lies on a sphere with finite non-null radius such that circles are mapped on circles, then (56) holds[12].

Sufficiency. If (56) holds, then (37) holds. It follows that (9) and (49) hold; therefore, as in the proof of Theorem 2, we obtain the system of equa-

(11) We have already indicated that if one non-null circle in D is mapped on a circle on S, then all circles in D are mapped on circles on S. See the last paragraph of the first part of the proof of Lemma 1.

(12) See the latter half of the proof of Theorem 2.

tions (52). Then, as before, it follows that the functions (6) either are the coordinate functions of a minimal surface in isothermic representation or they map D isothermically on a surface S that lies on a sphere with finite non-null radius[13]. Here we have made use of the results of the second part of the proof of Theorem 2. But in the first part of the proof of Theorem 2 we have shown that if S lies on a sphere of finite non-null radius, (37) implies that the first quadratic form of S has the representation (15); hence, by Lemma 1, all circles in D are mapped on circles on S.

COROLLARY 1. *If the functions* (6) *have continuous partial derivatives of the third order in a simply connected domain D, then a necessary and sufficient condition that*

$$\sum_{j=1}^{3}\left[\int_{C} x_j(u, v)dz\right]^2 = 0$$

hold for each circle C in D is that for each point (u_0, v_0) in D,

$$\sum_{j=1}^{3}\left[\int_{C_r} x_j(u, v)dz\right]^2 = o(r^8)$$

hold, where C_r is the circle in D with center (u_0, v_0) and radius r.

COROLLARY 2. *If the functions* (6) *have continuous partial derivatives of the third order in a simply connected domain D, then a necessary and sufficient condition that the functions* (6) *either map D isothermically on a minimal surface, or map D isothermically on a surface S that lies on a sphere with finite non-null radius such that circles are mapped on circles, is that for each point (u_0, v_0) in D,*

$$\sum_{j=1}^{3}\left[\int_{C_r} x_j(u, v)dz\right]^2 = o(r^8)$$

hold, where C_r is the circle in D with center at (u_0, v_0) and radius r.

4. CHARACTERIZATION OF ISOTHERMIC PLANE MAPS

LEMMA 2. *Let the function $g_1(z)$ be schlicht*[14], *and the function $g_2(z)$ analytic, in the circle $|z| < \rho$. If δ is an arbitrary positive number, $0<\delta<\rho$, then there exists a positive number $\epsilon = \epsilon(\delta)$ such that the function*

$$f(z) \equiv g_1(z) + \eta g_2(z) \tag{57}$$

is schlicht in the circle $|z| \leq \rho - \delta$, for each η such that $|\eta| < \epsilon$.

Proof. If the lemma does not hold, then there exists a positive number δ_0,

---

[13] If the sphere were either a point-sphere or a plane, then the functions (6) would form a triple of conjugate harmonic functions.

[14] The analytic function $g_1(z)$ is said to be *schlicht* in D if $g_1(z_1) = g_1(z_2)$ implies $z_1 = z_2$, where z_1 and z_2 are points of D.

two points, z_0' and z_0'', and a triple of sequences

$$\eta_1, \eta_2, \cdots,$$

$$z_1', z_2', \cdots, \qquad z_1'', z_2'', \cdots, \qquad z_n' \neq z_n'', \; n = 1, 2, \cdots,$$

with the properties

$$\lim_{n\to\infty} \eta_n = 0,$$

$$\lim_{n\to\infty} z_n' = z_0', \qquad |z_n'| \leqq \rho - \delta_0, \qquad n = 0, 1, 2, \cdots,$$

$$\lim_{n\to\infty} z_n'' = z_0'', \qquad |z_n''| \leqq \rho - \delta_0, \qquad n = 0, 1, 2, \cdots,$$

$$f_n(z_n') = f_n(z_n''), \qquad n = 1, 2, \cdots,$$

where

$$f_n(z) \equiv g_1(z) + \eta_n g_2(z), \qquad n = 1, 2, \cdots.$$

It follows immediately that

$$g_1(z_0') = g_1(z_0'').$$

But since $g_1(z)$ is *schlicht* in the circle $|z| < \rho$, it follows that $z_0' = z_0'' = z_0$. Moreover, we obtain

$$\lim_{n\to\infty} \frac{f_n(z_n') - f_n(z_n'')}{z_n' - z_n''} = \frac{dg_1(z_0)}{dz} = 0^{(15)}. \tag{58}$$

Since $|z_0| \leqq \rho - \delta_0$, (58) implies that $g_1(z)$ is not *schlicht* in the circle $|z| < \rho$. From this contradiction of the hypothesis, it follows that there exists an $\epsilon_0 > 0$ such that the function (57) is *schlicht*, for each η satisfying $|\eta| < \epsilon_0$, in $|z| \leqq \rho - \delta_0$.

THEOREM 4. *If the functions (6) have continuous partial derivatives of the third order in a simply connected domain D, then a necessary and sufficient condition that they map D isothermically on a plane surface is that for all closed rectifiable Jordan curves γ lying in D*

$$\sum_{j=1}^{3} \left[\int_\gamma x_j(u, v) dz \right]^2 = 0. \tag{59}$$

(15) Since $g_1(z)$ is analytic in the closed region defined by $|z| \leqq \rho - \delta_0$, for any arbitrary positive number δ' there exists a positive number ϵ', which is independent of z in the region $|z| \leqq \rho - \delta_0$, for which

$$\left| \frac{g_1(z') - g_1(z'')}{z' - z''} - \frac{d}{dz} g_1(z) \right| < \delta'$$

for all z', z'' and z in the region $|z| \leqq \rho - \delta_0$ that satisfy the inequalities $|z - z'| < \epsilon'$, $|z - z''| < \epsilon'$, $z' \neq z''$.

Necessity. If the functions (6) map D isothermically on a surface S that lies on a plane p, we make a rigid transformation in the x_1, x_2, x_3-space such that p coincides with the plane $x_3'=0$ and the positive normal at the image of an arbitrary fixed point (u_0, v_0) of D coincides with the positive x_3'-axis. Let this rigid transformation have the representation

$$(60) \qquad x_j' = \sum_{k=1}^{3} \lambda_{kj} x_k + a_j, \qquad j = 1, 2, 3,$$

where

$$(61) \qquad \begin{aligned} \sum_{k=1}^{3} \lambda_{sk}\lambda_{jk} &= 0, \quad s \neq j, \\ &= 1, \quad s = j, \qquad s, j = 1, 2, 3, \end{aligned}$$

and where the a_j, $j=1, 2, 3$, are real constants. Therefore the functions

$$(62) \qquad \begin{aligned} x_j'(u, v) &= \sum_{k=1}^{3} \lambda_{kj}[x_k(u, v) - x_k(u_0, v_0)], \qquad j = 1, 2, 3, \\ x_3'(u, v) &\equiv 0, \end{aligned}$$

map D isothermically on a plane surface, and hence, by the theorem of Weierstrass, they form a triple of conjugate harmonic functions. Since it follows from (8), (9), (61) and (62) that (3) holds for the functions $x_1'(u, v)$ and $x_2'(u, v)$, the functions $x_1'(u, v)$ and $x_2'(u, v)$ are a couple of conjugate harmonic functions. Hence by Cauchy's theorem, analogous to (5), and (62), it follows that

$$\sum_{j=1}^{3} \left[\int_\gamma x_j'(u, v) dz \right]^2 = 0,$$

which, with (4), (60) and (61), yields (59).

Sufficiency. If (59) holds for all closed rectifiable Jordan curves γ lying in D, then, by Theorem 3, the functions (6) either are the coordinate functions of a minimal surface in isothermic representation, or map D isothermically on a surface S that lies on a sphere of finite non-null radius such that circles are mapped on circles.

Part I. We first consider the case when the functions (6) are the coordinate functions of a minimal surface in isothermic representation. Let (u_0, v_0) be an arbitrary point in D and let the arbitrary fixed circle C_0 in D have its center at (u_0, v_0). If γ is an arbitrary closed rectifiable Jordan curve lying in C_0, which contains (u_0, v_0) in its interior, then the interior of γ can be mapped conformally on the interior of the circle $C: s^2+t^2=\rho^2$ in the s, t-plane such that the image of (u_0, v_0) is the center of C. The expansion of the inverse of this mapping function has the representation

$$z = h(w) = z_0 + \sum_{m=1}^{\infty} b_m w^m, \tag{63}$$

where

$$w = s + it, \qquad z_0 = u_0 + iv_0.$$

Since the functions (6) are the isothermic coordinate functions of a minimal surface, it follows, from the theorem of Weierstrass, that these functions may be written in the form

$$x_j(u, v) = \tfrac{1}{2}[\Phi_j(z) + \overline{\Phi_j(z)}], \qquad j = 1, 2, 3, \tag{64}$$

where $\Phi_j(z)$ is a function which is analytic in D and where $\overline{\Phi_j(z)}$ is its conjugate function. We may write

$$\Phi_j(z) = \sum_{n=0}^{\infty} a_{j,n}(z - z_0)^n, \qquad j = 1, 2, 3;$$

then

$$\overline{\Phi_j(z)} = \sum_{n=0}^{\infty} \bar{a}_{j,n}\overline{(z - z_0)}^n, \qquad j = 1, 2, 3,$$

where the series for $\Phi_j(z)$ and $\overline{\Phi_j(z)}$ are absolutely convergent in the interior of and on C_0. Therefore

$$\overline{\Phi_j(z)} = \bar{a}_{j,0} + \sum_{n=1}^{\infty} \bar{a}_{j,n} \sum_{m=1}^{\infty} P_{n,m}\bar{w}^m, \qquad j = 1, 2, 3, \tag{65}$$

where

$$\begin{aligned} P_{n,m} &= \sum_{k_l} \bar{b}_{k_1} \cdots \bar{b}_{k_n}, \quad n \leqq m, \ k_1 + \cdots + k_n = m, \ k_l \geqq 1, \ l = 1, 2, \cdots, n, \\ P_{n,m} &= 0, \qquad n > m. \end{aligned} \tag{66}$$

If C_R is the circle concentric with C and of radius R, $0 < R < \rho$, then the function (63) maps the interior of C_R conformally on the interior of a closed rectifiable Jordan curve γ_R that lies in the interior of γ. After we have set $w = Re^{i\phi}$ in (63) and (65), it follows from (4), (63), (64), (65) and (66) that

$$\sum_{j=1}^{3}\left[\int_{\gamma_R} x_j(u, v)dz\right]^2 = -\pi^2 \sum_{t=2}^{\infty}\sum_{m=1}^{t-1}\sum_{n=1}^{m}\sum_{s=1}^{t-m} A_{n,s}B_{n,m}B_{s,t-m}R^{2t},$$

which, with (59), implies

$$\sum_{t=2}^{\infty}\sum_{m=1}^{t-1}\sum_{n=1}^{m}\sum_{s=1}^{t-m} A_{n,s}B_{n,m}B_{s,t-m}R^{2t} = 0, \tag{67}$$

where

$$B_{n,m} = mb_m P_{n,m}, \qquad n, m = 1, 2, \cdots, \tag{68}$$

and where

$$A_{n,s} = A_{s,n} = \sum_{j=1}^{3} \bar{a}_{j,n}\bar{a}_{j,s}, \qquad n, s = 1, 2, \cdots. \tag{69}$$

But the circle C_R is an arbitrary circle concentric with and interior to the circle C; therefore the relation (67) is independent of R, $0<R<\rho$. Hence

$$\sum_{m=1}^{t-1}\sum_{n=1}^{m}\sum_{s=1}^{t-m} A_{n,s}B_{n,m}B_{s,t-m} = 0, \qquad t = 2, 3, \cdots. \tag{70}$$

From (70) we obtain, by an induction,

$$A_{n,s} = 0, \qquad n, s = 1, 2, \cdots. \tag{71}$$

For, when $t=2$ in (70) we obtain, by (66) and (68),

$$A_{1,1}b_1^2 b_1^2 = 0. \tag{72}$$

But the function (63) maps the interior of C on the interior of γ in a one-to-one manner[16]. Therefore $b_1 \neq 0$. Hence it follows from (72) that

$$A_{1,1} = 0. \tag{73}$$

Now suppose that

$$A_{n,s} = 0, \qquad 2 \leqq n + s \leqq p - 1,\ p \geqq 3, \tag{74}$$

where n, s and p are positive integers. We shall show that

$$A_{n,s} = 0, \qquad n + s = p. \tag{75}$$

For $t=p$, (70) yields, by (66), (68), (74) and the fact that $b_1 \neq 0$,

$$\sum_{k=1}^{p-1} k(p-k)b_k b_{p-k} A_{k,p-k} = 0. \tag{76}$$

By Lemma 2, for a fixed R, $0<R<\rho$, the functions

$$G_q(w) = h(w) + \eta_q w^q = z_0 + \sum_{m=1}^{\infty} b_m w^m + \eta_q w^q, \quad q = 1, 2, \cdots, p-1, \tag{77}$$

are *schlicht* in the circle $|w| \leqq R$, provided

$$|\eta_q| < \epsilon_q, \qquad q = 1, 2, \cdots, p-1,$$

where ϵ_q is a positive constant whose existence was established in Lemma 2.

---

(16) This is a consequence of Darboux's theorem. See W. F. Osgood, *Functions of a Complex Variable*, p. 167.

We further restrict η_q to satisfy the relation

$$R^q|\eta_q| < \delta,$$

for $q=1, 2, \cdots, p-1$, where δ is the distance between γ_R and C_0, in order that the map $\gamma_{q,R}$ of C_R by (77) shall lie inside C_0. Moreover, (u_0, v_0) is inside $\gamma_{q,R}$. Hence if

$$0 < |\eta_q| < d_q = \min\left[\epsilon_q, \frac{\delta}{R^q}\right], \qquad q = 1, 2, \cdots, p-1, \tag{78}$$

then we obtain the following result, which is analogous to (76), for the functions (77):

$$\sum_{k=1}^{p-1} k(p-k)b_k b_{p-k} A_{k,p-k} + 2q(p-q)b_{p-q}\eta_q A_{q,p-q} = 0, \tag{79a}$$

$$\sum_{k=1}^{p-1} k(p-k)b_k b_{p-k} A_{k,p-k} + q^2(\eta_q^2 + 2\eta_q b_q)A_{q,q} = 0, \tag{79b}$$

where (79a) holds for all q, $1 \leqq q \leqq p-1$, except $q=p/2$, and where (79b) holds for $q=p/2$[17]. From (76) and (79) we obtain

$$2q(p-q)b_{p-q}\eta_q A_{q,p-q} = 0, \tag{80a}$$

$$q^2(\eta_q^2 + 2\eta_q b_q)A_{q,q} = 0, \tag{80b}$$

where (80a) holds for all q, $1 \leqq q \leqq p-1$, except $q=p/2$, and where (80b) holds for $q=p/2$.

Since η_q is an arbitrary constant which is subject only to the restriction (78), it follows from the relation (80b) that

$$A_{q,q} = 0, \qquad q = p/2, \tag{81}$$

provided p is even. If q_1 is a fixed positive integer, $1 \leqq q_1 \leqq p-1$, and $q_1 \neq p/2$, and if $b_{p-q_1} \neq 0$, then it follows from (80a) that

$$A_{q_1,p-q_1} = 0. \tag{82}$$

If q_2 is a fixed positive integer, $1 \leqq q_2 \leqq p-1$ and $q_2 \neq p/2$, and if $b_{p-q_2} = 0$, then we consider the function

$$h_1(w) = h(w) + \eta'_{p-q_2} w^{p-q_2}, \tag{83}$$

where η'_{p-q_2} is a fixed constant,

$$0 < |\eta'_{p-q_2}| < d_{p-q_2}.$$

---

[17] When we state "except for $q=p/2$," we mean that if p is an even integer, then q does not take on the integral value $p/2$.

The function (83) is *schlicht* in the circle $|w| \leq R$([18]), and maps C_R on a closed rectifiable Jordan γ_R' that lies inside C_0. Moreover, (u_0, v_0) is inside γ_R'. From (63) and (83) we obtain

$$(84) \qquad h_1(w) = z_0 + \sum_{m=1}^{\infty} c_m w^m,$$

where

$$c_m = b_m, \quad m \neq p - q_2; \qquad c_{p-q_2} = \eta'_{p-q_2}.$$

If we apply (84) as we have applied (63) in the earlier part of this proof of Theorem 4, then we obtain the following result analogous to (80a):

$$(85) \qquad 2q_2(p - q_2)\eta'_{p-q_2}\eta''_{q_2}A_{q_2,p-q_2} = 0,$$

where η''_{q_2} is an arbitrary constant,

$$(86) \qquad 0 < |\eta''_{q_2}| < d'_{q_2}.$$

Here the constant d'_{q_2} is in the same relation to $h_1(w)$ as d_{q_2} is to $h(w)$. Since $\eta'_{p-q_2} \neq 0$ and $q_2 \neq p/2$, it follows from (85) and (86) that

$$(87) \qquad A_{q_2,p-q_2} = 0.$$

From (81), (82) and (87) it follows that (75) holds. Since (73) holds, our induction is now complete.

From (69), (71) and the substitution

$$a_{j,m} = \alpha_{j,m} - i\beta_{j,m}, \qquad j = 1, 2, 3, \; m = 0, 1, \cdots,$$

we get the following relations:

$$(88) \qquad \begin{aligned} &\sum_{j=1}^{3} (\alpha_{j,n}\alpha_{j,s} - \beta_{j,n}\beta_{j,s}) = 0, \\ &\sum_{j=1}^{3} (\alpha_{j,n}\beta_{j,s} + \alpha_{j,s}\beta_{j,n}) = 0, \qquad n, s \geq 1 \text{([19])}; \end{aligned}$$

here $\alpha_{j,0}$, $\alpha_{j,m}$ and $\beta_{j,m}$, $m = 1, 2, \cdots$, are the Fourier coefficients in the expansion of $x_j(u, v)$, $j = 1, 2, 3$, about the point (u_0, v_0).

We transform the axes in the x_1, x_2, x_3-space such that the new origin is at the image of (u_0, v_0), the plane $x_3' = 0$ is tangent to the surface there, and the positive normal to the surface there coincides with the positive x_3'-axis. This transformation is given by (60) and (61), where

([18]) This holds because $|\eta'_{p-q_2}| < d_{p-q_2}$, where d_{p-q_2} was defined in (78).

([19]) Compare with J. W. Hahn and E. F. Beckenbach, *Triples of conjugate harmonic functions and minimal surfaces*, Duke Mathematical Journal, vol. 2 (1936), Lemma 1, p. 699, and footnote, p. 700.

$$a_j = -\sum_{k=1}^{3} \lambda_{kj} x_k(u_0, v_0), \qquad j = 1, 2, 3.$$

The relations (88) are invariant under this rigid transformation. For, since the functions (6) form a triple of conjugate harmonic functions, it follows that the new coordinate functions,

$$x_j'(u, v) = \sum_{k=1}^{3} \lambda_{kj}[x_k(u, v) - x_k(u_0, v_0)], \qquad j = 1, 2, 3,$$

form a triple of conjugate harmonic functions. Hence we may write, in terms of polar coordinates,

$$x_j'(u, v) = \alpha_{j,0}' + \sum_{m=1}^{\infty} r^m(\alpha_{j,m}' \cos m\theta + \beta_{j,m}' \sin m\theta), \qquad j = 1, 2, 3, \tag{89}$$

where

$$\alpha_{j,m}' = \sum_{k=1}^{3} \lambda_{kj}\alpha_{k,m}, \quad \beta_{j,m}' = \sum_{k=1}^{3} \lambda_{kj}\beta_{k,m}, \qquad m = 1, 2, \cdots, j = 1, 2, 3. \tag{90}$$

From (61), (88) and (90) we obtain

$$\begin{aligned} &\sum_{j=1}^{3} (\alpha_{j,n}'\alpha_{j,s}' - \beta_{j,n}'\beta_{j,s}') = 0, \\ &\sum_{j=1}^{3} (\alpha_{j,n}'\beta_{j,s}' + \alpha_{j,s}'\beta_{j,n}') = 0, \qquad n, s \geqq 1. \end{aligned} \tag{91}$$

Therefore the relations (88) are invariant under a rigid transformation.

To prove that the functions (6) define a plane surface, it is sufficient to show

$$\alpha_{3,0}' = \alpha_{3,m}' = \beta_{3,m}' = 0, \qquad m = 1, 2, \cdots. \tag{92}$$

Since $x_j'(u_0, v_0) = 0, j = 1, 2, 3$, it follows that

$$\alpha_{j,0}' = 0, \qquad j = 1, 2, 3. \tag{93}$$

Let t be the positive integer for which

$$\sum_{j=1}^{3} \alpha_{j,m}'^2 = 0, \qquad m = 0, 1, \cdots, t - 1, \tag{94a}$$

$$\sum_{j=1}^{3} \alpha_{j,t}'^2 \neq 0[20]. \tag{94b}$$

---

[20] If $\sum_{j=1}^{3}\alpha_{j,m}'^2 = 0$ for all positive m, then it follows, from the first relation in (91), that $\sum_{j=1}^{3}\beta_{j,m}'^2 = 0$ for all positive m; and therefore (92) holds.

From (89) we obtain

$$\alpha'_{2,t}\beta'_{3,t} - \alpha'_{3,t}\beta'_{2,t} = 0, \tag{95a}$$

$$\alpha'_{3,t}\beta'_{1,t} - \alpha'_{1,t}\beta'_{3,t} = 0, \tag{95b}$$

$$\alpha'^2_{1,t} + \alpha'^2_{2,t} + \alpha'^2_{3,t} = \alpha'_{1,t}\beta'_{2,t} - \alpha'_{2,t}\beta'_{1,t}\text{[21]}, \tag{95c}$$

since the positive normal to S at the point $(0, 0, 0)$ coincides with the positive x_3'-axis. Equations (95a) and (95b) are linear and homogeneous in $\alpha'_{3,t}$ and $\beta'_{3,t}$; since by (94b) and (95c) their determinant of coefficients is not zero, it follows that

$$\alpha'_{3,t} = \beta'_{3,t} = 0. \tag{96}$$

From (91), for $n=s=t$, (94b) and (96) we obtain

$$\begin{aligned} \alpha'^2_{1,t} + \alpha'^2_{2,t} &= \beta'^2_{1,t} + \beta'^2_{2,t} \neq 0, \\ \alpha'_{1,t}\beta'_{1,t} + \alpha'_{2,t}\beta'_{2,t} &= 0, \end{aligned} \tag{97}$$

which, with (95c) and (96), imply

$$\alpha'_{1,t} = \beta'_{2,t}, \qquad \alpha'_{2,t} = -\beta'_{1,t}. \tag{98}$$

From (91), for $s=t\leqq n$, (96) and (98) we obtain

$$\begin{aligned} \alpha'_{1,n}\alpha'_{1,t} + \alpha'_{2,n}\alpha'_{2,t} + \alpha'_{2,t}\beta'_{1,n} - \alpha'_{1,t}\beta'_{2,n} &= 0, \\ -\alpha'_{1,n}\alpha'_{2,t} + \alpha'_{2,n}\alpha'_{1,t} + \alpha'_{2,t}\beta'_{2,n} + \alpha'_{1,t}\beta'_{1,n} &= 0, \qquad n \geqq t. \end{aligned}$$

If we eliminate first $\alpha'_{2,n}$ and then $\beta'_{2,n}$ from these last two relations, then we obtain two other relations which, with the first relation in (97), imply

$$\alpha'_{1,n} = \beta'_{2,n}, \qquad \alpha'_{2,n} = -\beta'_{1,n}, \qquad n \geqq t. \tag{99}$$

From (91), for $n=s\geqq t$, and (99) we obtain

$$\alpha'_{3,n}\beta'_{3,n} = 0, \qquad \alpha'^2_{3,n} = \beta'^2_{3,n}, \qquad n \geqq t,$$

from which it follows that

$$\alpha'_{3,n} = \beta'_{3,n} = 0, \qquad n \geqq t. \tag{100}$$

From (91), for $n=s<t$, (94a) and (100), we obtain (92).

---

[21] If $z-z_0=r(\cos\theta+i\sin\theta)$, then for r sufficiently small the components of the unit normal vector to the surface are $\zeta_j=\zeta_j(u, v)$, $j=1, 2, 3$, $\zeta_j(u, v)=[(\alpha'_{k,t}\beta'_{l,t}-\alpha'_{l,t}\beta'_{k,t})/\sum_{j=1}^{3}\alpha'^2_{j,t}]+O(r)$, where $O(r)$ denotes a quantity $\psi(r)$ (not always the same quantity) such that $|\psi(r)/r|$ is bounded. Note that $\zeta_1=0$, $\zeta_2=0$, $\zeta_3=1$ at $(0, 0, 0)$.

COROLLARY. *If the functions* (6) *are harmonic in a simply connected domain* D, *such that their Fourier expansions about the fixed point* (u_0, v_0) *of* D *are*

$$x_j(u, v) = \alpha_{j,0} + \sum_{m=1}^{\infty} r^m(\alpha_{j,m} \cos m\theta + \beta_{j,m} \sin m\theta), \qquad j = 1, 2, 3,$$

and if

$$\sum_{j=1}^{3} (\alpha_{j,n}\alpha_{j,s} - \beta_{j,n}\beta_{j,s}) = 0, \qquad \sum_{j=1}^{3} (\alpha_{j,n}\beta_{j,s} + \alpha_{j,s}\beta_{j,n}) = 0, \qquad n, s \geqq 1,$$

then the functions (6) *are the isothermic coordinate functions of a plane surface.*

Part II. We now consider the case when the functions (6) map D isothermically on a surface S that lies on a sphere $\mathfrak{S}$, whose radius a is finite and non-null, such that circles are mapped on circles. Since the functions (6) have the representation (18), it follows that they may be continued isothermically to map the whole u, v-plane isothermically on $\mathfrak{S}$ such that circles are mapped on circles. Let the point P on $\mathfrak{S}$ correspond to the point $z=\infty$, and let p be the equatorial plane corresponding to P as a pole. Then the stereographic projection of $\mathfrak{S}$ on p, with P as pole, induces an isothermic map of the u, v-plane on p such that circles are mapped on circles, and such that the point at infinity in the u, v-plane corresponds to the point at infinity in the plane p. Let us take the center M of $\mathfrak{S}$ as the origin in a system of coordinates on p, such that the positive s-axis, the positive t-axis, and the ray MP, in that order, have the same disposition as the coordinate axes (x_1, x_2, x_3). The mapping function must have one of the following representations:

$$w = s + it = f(z) = \alpha z + \beta, \tag{101a}$$

or

$$w = s + it = f(z) = \alpha \bar{z} + \beta, \qquad \alpha \neq 0, \tag{101b}$$

where (101a) holds if the map of the u, v-plane on p is directly conformal, and where (101b) holds if the map is inversely conformal.

Consider the system of axes (x_1', x_2', x_3') in space, with origin at M, and with the positive x_1'-, x_2'- and x_3'-axes having the directions of the positive s-axis, the positive t-axis and the ray MP respectively. There exists a rigid transformation in the x_1, x_2, x_3-space which carries the origin into the point M, and which carries the positive x_1-, x_2- and x_3-axes into the positive x_1'-, x_2'- and x_3'-axes respectively; this transformation has the form (60), where (61) holds. Let the new coordinate functions of S be

$$x_j' = x_j'(u, v), \qquad j = 1, 2, 3. \tag{102}$$

Since the functions (6) map circles on circles, the functions (102) map circles on circles; therefore, as noted in Lemma 1, the functions (102) have the form

(18) where $f(z)$ is given by (101). Hence, since $\alpha \neq 0$ in (101), we have, for the arbitrary closed rectifiable Jordan curve γ in D,

$$\sum_{j=1}^{3}\left[\int_{\gamma} x_j'(u, v)dz\right]^2 = \frac{4a^4}{\alpha^2}\sum_{j=1}^{3}\left[\int_{\gamma'} \phi_j(s, t)dw\right]^2, \tag{103}$$

where γ' is the image of γ on the plane p and where

$$\phi_1(s, t) = \frac{s}{a^2 + s^2 + t^2}, \quad \phi_2(s, t) = \frac{t}{a^2 + s^2 + t^2}, \quad \phi_3(s, t) = \frac{a}{a^2 + s^2 + t^2}.$$

As it previously has been shown[22], (59) is invariant under a rigid transformation; hence, since the radius a of $\mathfrak{S}$ is different from zero, it follows from (59) and (103) that

$$\sum_{j=1}^{3}\left[\int_{\gamma'} \phi_j(s, t)dw\right]^2 = 0 \tag{104}$$

holds. Since γ is an arbitrary closed rectifiable Jordan curve in D, it follows that (104) holds for each closed rectifiable Jordan curve γ' in D', where D' is the map on the plane p by (101).

Without any loss of generality, we may assume the line $s=0$ passes through D'. Then there exists a closed rectifiable Jordan curve Γ in D' with the following description:

(1) the vertices of Γ have the following polar coordinates,

$$A_1:\ (r, \tau), \qquad A_2:\ (r, -\tau),$$
$$A_3:\ (r + \omega, -\tau), \quad A_4:\ (r + \omega, \tau), \qquad 0 < \tau < \pi/2, 0 < r, 0 < \omega;$$

(2) Γ is composed of two arcs of circles and two straight-line segments: arc A_1A_2 is an arc of the circle $s^2+t^2=r^2$, arc A_3A_4 is an arc of the circle $s^2+t^2=(r+\omega)^2$, each arc subtending an angle 2τ at the origin, and A_2A_3 and A_4A_1 are on rays through the origin.

For the closed rectifiable Jordan curve Γ we obtain

$$\sum_{j=1}^{3}\left[\int_{\Gamma} \phi_j(s, t)dw\right]^2 = -\,\xi\,\chi_1(r, \omega)\sin\xi - 4\,\chi_2^2(r, \omega)\sin^2\frac{\xi}{2}, \qquad \xi = 2\tau,$$

where

$$\chi_1(r, \omega) \equiv \left[\frac{(r+\omega)^2}{a^2 + (r+\omega)^2} - \frac{r^2}{a^2 + r^2}\right]\left[\frac{(r+\omega)^2}{a^2 + (r+\omega)^2} - \frac{r^2}{a^2 + r^2} + \log\frac{a^2 + r^2}{a^2 + (r+\omega)^2}\right],$$

(22) See the first part of this proof of Theorem 4.

$$\chi_2(r, \omega) \equiv \left[a\left(\frac{r+\omega}{a^2+(r+\omega)^2} - \frac{r}{a^2+r^2}\right) + \text{arc tan}\,\frac{r}{a} - \text{arc tan}\,\frac{r+\omega}{a}\right].$$

Since

$$\frac{\partial\chi_2}{\partial\omega} = -\frac{2a(r+\omega)^2}{[a^2+(r+\omega)^2]^2} \not\equiv 0,$$

it follows that $\chi_2(r, \omega)\not\equiv 0$. Fix r and ω so that $\chi_2(r, \omega)\neq 0$. Since $\xi \sin \xi$ and $4 \sin^2 \xi/2$ are not proportional,

$$\xi \sin \xi = \xi^2 - \frac{\xi^4}{6} + \cdots, \qquad 4 \sin^2 \frac{\xi}{2} = \xi^2 - \frac{\xi^4}{12} + \cdots,$$

there exists a ξ such that

$$\sum_{j=1}^{3}\left[\int_\Gamma \phi_j(s, t)dw\right]^2 \neq 0. \tag{105}$$

Now (105) is a contradiction of (104). Therefore the functions (6) do not map D on a surface that lies on a sphere of finite non-null radius.

THE RICE INSTITUTE,
HOUSTON, TEXAS.

FAMILIES OF CURVES CONFORMALLY EQUIVALENT TO CIRCLES

BY

EDWARD KASNER AND JOHN DE CICCO

1. **Introduction.** In this paper, we shall study three-parameter families of curves conformally equivalent to the totality of circles. We obtain the analytic form and several complete geometric characterizations of any such family, which we shall call an Ω *family*.

In minimal coordinates (u, v) the differential equation of such a family is of the form

$$2\frac{dv}{du}\frac{d^3v}{du^3} - 3\left(\frac{d^2v}{du^2}\right)^2 = 2\left(\frac{dv}{du}\right)^2\left[A(u) - \left(\frac{dv}{du}\right)^2 B(v)\right].$$

Our fundamental result is that any Ω family may be characterized among all three-parameter families of curves by the following three properties: *Property* I. The locus of the foci of the osculating parabolas of the ∞^1 curves which contain a given lineal element E is a lemniscate L with E as one of the two orthogonal tangent elements at the node of L. *Property* II. As the direction of E is rotated about its point P, the locus of the centers of the orthogonal pairs of circles defining the ∞^1 focal lemniscates is an equilateral hyperbola H with the center of H at P. *Property* III. The foci of the equilateral hyperbolas are connected to the point P by a direct conformal transformation.

Another complete geometric characterization of an Ω family of curves is the following. *Property* I′. The envelope of the directrices of the osculating parabolas of the ∞^1 curves which contain a given lineal element E is an equilateral hyperbola H with the point of E as the center of H and the line of E as one of the asymptotes of H. *Property* II′. The locus of the foci of the directorial equilateral hyperbolas is an equilateral hyperbola H' with its center at P. *Property* III′. The four foci of the equilateral hyperbola H' of Property II′ are related to the point P by a direct conformal transformation.

A reciprocity relation appears in these two sets of geometric characterizations. But this reciprocal relation is by no means self-evident. For although the two characterizations are roughly dual, separate proofs are required. The lemniscate and the equilateral hyperbola of Properties I and I′ are equivalent under inversion, whereas the equilateral hyperbolas of Properties II and II′ are equivalent under a similitude.

In connection with this duality, the following results may be noted. The ∞^1 focal lemniscates of Property I, constructed at the point P, all pass

Presented to the Society, April 27, 1940; received by the editors April 10, 1940.

through the vertices of a minimal quadrilateral with center at P. The ∞^1 directorial equilateral hyperbolas of Property I′, constructed at P, are tangent to the sides of a quadrilateral whose diagonals are the minimal lines through P.

In the latter part of our paper, we shall give other geometric properties of an Ω family of curves. The rate κ' of variation of the curvature with respect to the arc length is the same for all the curves of our family which contain a lineal element E. The resulting ∞^1 rates of variation of curvature at a point P are connected by an analogue of Meusnier's theorem. We note that our problem may be considered to be a generalization of the family of curves conformally equivalent to the ∞^2 straight lines. This simple type has been considered elsewhere. Relations to dynamical and natural families are of interest.

2. The differential equation of any Ω family. For the analytic work, we shall find it convenient to use the minimal coordinates (u, v) instead of the ordinary rectangular cartesian coordinates (x, y). These are connected by the relations

$$u = x + iy, \qquad v = x - iy. \tag{1}$$

An Ω *family* of curves consists of a three-parameter set which is equivalent to the ∞^3 circles under a given conformal transformation. We proceed to derive the differential equation of any such family. However, before this can be accomplished, it is necessary to discuss some preliminary material.

In minimal coordinates (u, v), any direct conformal transformation is

$$U = \phi(u), \qquad V = \psi(v), \tag{2}$$

where the functions are of course analytic. A reverse conformal transformation may be expressed as the product of a direct conformal transformation by the reflection through the x-axis: $U=v$, $V=u$. Upon extending this conformal transformation three times, we find

$$\begin{aligned}
P &= \frac{\psi_v}{\phi_u} p,\\
P' &= \frac{\psi_v}{\phi_u^2} p' + \frac{\psi_{vv}}{\phi_u^2} p^2 - \frac{\psi_v \phi_{uu}}{\phi_u^3} p,\\
P'' &= \frac{\psi_v}{\phi_u^3} p'' + \frac{3\psi_{vv}}{\phi_u^3} pp' - \frac{3\psi_v \phi_{uu}}{\phi_u^4} p' + \frac{\psi_{vvv}}{\phi_u^3} p^3 - \frac{3\psi_{vv}\phi_{uu}}{\phi_u^4} p^2\\
&\quad + \frac{\psi_v}{\phi_u^5}(-\phi_u \phi_{uuu} + 3\phi_{uu}^2) p,
\end{aligned} \tag{3}$$

where $p = dv/du$ and $P = dV/dU$.

The ∞^3 circles of the plane are represented in minimal coordinates by the ∞^3 conics which pass through the two fixed points at infinity, given in homo-

geneous minimal coordinates by (0, 1, 0) and (1, 0, 0). Therefore the equation of any circle is

$$auv + bu + cv + d = 0. \tag{4}$$

From this, the differential equation of all ∞^3 circles is found to be

$$2pp'' - 3p'^2 = 0. \tag{5}$$

The Moebius group of the complex plane is the group of all point transformations which preserve the entire family of ∞^3 circles. In minimal coordinates, any direct Moebius transformation may be written as

$$U = \frac{au + b}{cu + d}, \qquad V = \frac{a'v + b'}{c'v + d'}. \tag{6}$$

A reverse Moebius transformation is the product of a direct Moebius transformation by the reflection through the x-axis. A Moebius transformation is said to be *real* if the coefficients of one of the above equations are the respective conjugates of those of the other. In the complex cartesian plane, there are $2\infty^6$ Moebius transformations, of which $2\infty^3$ are real.

Upon applying the conformal transformation (2) to the circles of the (u, v) plane, the differential equation (5) expressed in capital letters is carried into the differential equation

$$2pp'' - 3p'^2 = 2p^2(A - Bp^2) \tag{7}$$

where

$$A = \frac{2\phi_u\phi_{uuu} - 3\phi_{uu}^2}{2\phi_u^2}, \qquad B = \frac{2\psi_v\psi_{vvv} - 3\psi_{vv}^2}{2\psi_v^2}. \tag{8}$$

From this, it is found that A may be any function of u only and B may be any function of v only. (All our functions in this paper are assumed to be analytic.)

THEOREM 1. *A differential equation of the third order represents an Ω family of curves if and only if it is of the form (where $p = v' = dv/du$)*

$$2pp'' - 3p'^2 = 2p^2(A - Bp^2), \tag{9}$$

where A is an arbitrary function of u and B is an arbitrary function of v.

By the preceding differential equation, we observe that any Ω family is uniquely determined by the functions A and B. Therefore Ω is a function of A and B only and we write $\Omega = \Omega(A, B)$.

If $\phi(u)$ and $\psi(v)$ are any two functions which satisfy (8), then the integral curves of (9) are the transforms under the conformal transformation (2) of the ∞^3 circles of the (U, V) plane. Therefore the curves of our Ω family are

$$a\phi\psi + b\phi + c\psi + d = 0. \tag{10}$$

Any Ω family of curves is thus a special type of linear families of curves. Of course, the differential equation (9) could have been obtained as a result of eliminating the arbitrary constants from the above equation by differentiation.

By (10), it is obvious that under a conformal transformation any Ω family is converted in general into some other Ω family. The group of transformations which preserve a given $\Omega(A, B)$ family of curves is a mixed six-parameter group $G_6'[\Omega(A, B)]$ isomorphic with the Moebius group of circular transformations. Any transformation Σ of this group is of the form TMT^{-1} where T is a definite conformal transformation which converts the ∞^3 circles into our Ω-family and M is any Moebius transformation. This group is generated by conformal symmetries (Schwarzian reflections) with respect to the curves of the Ω family.

Because of this isomorphism with the Moebius group, we may make the following observations(1).

THEOREM 2. *A nonconformal transformation converts at most* $2\infty^2$ *curves of the given* Ω *family into curves of the same family. A conformal transformation, not of the group* $G_6'[\Omega(A, B)]$, *converts at most* $2\infty^1$ *curves of the given* Ω *family into curves of the same family.*

From this theorem, we may conclude that a point transformation which converts $3\infty^2$ curves of a given Ω family into curves of the same family must belong to the group $G_6'[\Omega(A, B)]$. Similarly any conformal transformation which carries $3\infty^1$ curves of a given Ω family into curves of the same family must belong to the group $G_6'[\Omega(A, B)]$.

Let T and R be respectively a definite conformal transformation and any other transformation both of which carry the ∞^3 circles into a given $\Omega(A, B)$ family of curves. Then obviously $R = TM$, where M is a Moebius transformation.

Theorem 1 gives us an analytic characterization of any Ω family. In the remainder of our work, we shall give geometric characterizations of any such family. For this, we shall suppose that A and B of our family are each *not* identically zero. Any such family may be obtained by applying to the ∞^3 circles any conformal transformation written in the form (2) where ϕ is not a linear fractional function of u only and ψ is not a linear fractional function of v only.

3. **The osculating parabolas of any three-parameter family of curves.** Just as a set of values for (u, v, p, p'), that is, a differential element of the second order, is pictured most simply by means of the corresponding circle of curvature, so a differential element of the third order, defined by (u, v, p, p', p''),

(1) Kasner and De Cicco, *Characterization of the Moebius group of circular transformations*, Proceedings of the National Academy of Sciences, vol. 25 (1939), pp. 209–213.

may be pictured by the unique osculating parabola. We shall collect here the general formulas to be used in the subsequent discussion(²).

The equation of any parabola in minimal coordinates is

$$(11)\qquad (ku - lv)^2 + 2k(m + 2l\beta)u + 2l(m + 2k\alpha)v + (m^2 - 4kl\alpha\beta) = 0,$$

where (α, β) are the minimal coordinates of the focus and $ku+lv+m=0$ is the equation of the directrix. The unique osculating parabola of the differential element of the third order (u, v, p, p', p'') must have the parameters

$$(12)\qquad \begin{gathered} k = pp'' - 3p'^2, \qquad l = p'', \qquad m = -3pp' - ku - lv, \\ \alpha = u + \frac{3p'}{2p''}, \qquad \beta = v + \frac{3p^2p'}{2(pp'' - 3p'^2)}. \end{gathered}$$

Solving the last two equations for p' and p'', we find

$$(12')\qquad p' = \frac{p}{2}\left[\frac{1}{\alpha - u} - \frac{p}{\beta - v}\right], \qquad p'' = \frac{3p}{4(\alpha - u)}\left[\frac{1}{\alpha - u} - \frac{p}{\beta - v}\right].$$

Consider now any triply infinite system of curves, defined by a differential equation of third order

$$(13)\qquad p'' = f(u, v, p, p').$$

Through a given point in a given direction there pass ∞^1 curves of the system. Each of these has a definite osculating parabola at the given point. The locus of the foci of these parabolas is termed the *focal curve* and the envelope of the directrices is called the *directorial curve*. Thus to each lineal element $E(u, v, p)$ of the plane there corresponds a definite focal curve and a definite directorial curve.

The form of the focal curve depends, of course, upon the form of the differential equation. Since (u, v, p) are fixed, p'' is a certain function of p'. Substituting this in (12) and eliminating p', we obtain the finite equation of the required locus.

To obtain the finite form of the directorial curve, we proceed as follows. Upon writing the equation of the directrices of the osculating parabolas and differentiating it partially with respect to p', we obtain

$$(14)\qquad \begin{gathered} (pf - 3p'^2)(\alpha - u) + f(\beta - v) = 3pp', \\ (pf_{p'} - 6p')(\alpha - u) + f_{p'}(\beta - v) = 3p. \end{gathered}$$

Solving these for α and β, we find

(²) Kasner, *The trajectories of dynamics*, these Transactions, vol. 7 (1906), pp. 401–424. Also see the Colloquium volume by Kasner, *Differential-Geometric Aspects of Dynamics*, 1913; second edition, 1924. Recent discussion and extensions have been given by Moissiev, Fialkow, and MacColl.

$$(15)\qquad \begin{aligned} &\alpha = u + \frac{p(p'f_{p'} - f)}{p'(-p'f_{p'} + 2f)}, \qquad \beta = v + \frac{p(-pp'f_{p'} + pf + 3p'^2)}{p'(-p'f_{p'} + 2f)}, \\ &(\beta - v)^2 - p^2(\alpha - u)^2 = \frac{6p^3(-p'f_{p'} + f + 3p'^2/2p)}{(-p'f_{p'} + 2f)^2}. \end{aligned}$$

The third of these is written for later purposes. The elimination of p' from the first two of these equations gives the implicit form of our directorial curve.

4. **Characterization by focal curves.** We wish to consider the three-parameter families of curves such that the corresponding focal curve of any lineal element $E(u, v, p)$ of the plane is a lemniscate L with E as one of the two orthogonal tangent elements at the node of L. Before proceeding with this discussion, it is advisable to give a geometric construction of any lemniscate. Let C and C' be any two equal circles which are orthogonal at a point O. Draw any line l through O and let l intersect C and C' in the two points Q and Q' respectively. Let P be any point on l such that the distance OP is the mean proportion between the distances OQ and OQ'. (We note that the square of OP can have two values, one the negative of the other. But if we are careful to choose the signs of OP and OQ (OQ') such that the product of OP by QP $(Q'P)$ is equal to the square of the tangent from P to C (C'), then the square of OP can have only one value.) The set of all such points P forms the lemniscate of Bernoulli.

Let the minimal coordinates of the point O and the centers of C and C' be respectively (u, v), (a, b), $[u+i(a-u), v-i(b-v)]$. The equations of C and C' are respectively

$$(16)\qquad \begin{aligned} C:&\quad (\alpha - u)(\beta - v) = (a - u)(\beta - v) + (b - v)(\alpha - u), \\ C':&\quad (\alpha - u)(\beta - v) = i(a - u)(\beta - v) - i(b - v)(\alpha - u), \end{aligned}$$

where (α, β) are the running minimal coordinates of the points of C or C'. The equation of our lemniscate L is

$$(17)\qquad (\alpha - u)^2(\beta - v)^2 = i(a - u)^2(\beta - v)^2 - i(b - v)^2(\alpha - u)^2.$$

There are ∞^4 lemniscates in the complex plane. The node of our lemniscate L consists of the two orthogonal tangent elements of C and C' at the point O. Note that there are four pairs of orthogonal circles which define the same lemniscate L.

Upon substituting the values of α and β as given by (12) into (17) and noting that

$$(18)\qquad (b - v)^2/(a - u)^2 = p^2,$$

we obtain the following proposition.

THEOREM 3. *Property* I. *A three-parameter family of curves possesses the*

property that the corresponding focal curve of any lineal element E is a lemniscate L with E as one of the two orthogonal tangent elements at the node of L if and only if its differential equation is of the form

$$2pp'' - 3p'^2 = \lambda(u, v, p). \tag{19}$$

The lemniscates of our family of curves (19) are given by the equation

$$4\lambda(\alpha - u)^2(\beta - v)^2 = 3p^2[(\beta - v)^2 - p^2(\alpha - u)^2]. \tag{20}$$

From the above equation, we find that, if (a, b) is the center of any one circle of the four orthogonal pairs of circles which define any one of the lemniscates (20) then

$$i(a - u)^2 = 3p^2/4\lambda, \qquad i(b - v)^2 = 3p^4/4\lambda. \tag{21}$$

Let us now consider a three-parameter family of curves with the Property I. Of course any such family is given by a differential equation of the form (19). The focal curve associated to any lineal element E of the plane is a lemniscate. If we keep the point P of E fixed and vary the direction of E, the centers of the orthogonal pairs of circles defining these lemniscates will describe a locus. We shall call this locus the *central curve of the focal lemniscates* associated with the point P. By (19) and (21), we obtain the following result.

THEOREM 4. *Property* II. *A three-parameter family of curves with the Property* I *will possess the property that the central curve of the focal lemniscates associated with any point P of the plane is an equilateral hyperbola (eccentricity $e = \pm 2^{1/2}$) with its center at P if and only if its differential equation is of the form*

$$2pp'' - 3p'^2 = 2p^2(A - Bp^2), \tag{22}$$

where A and B are arbitrary functions of u and v only.

The equilateral hyperbolas of our family of curves (22) are

$$iA(a - u)^2 - iB(b - v)^2 = \tfrac{3}{8}. \tag{23}$$

The four foci of this equilateral hyperbola are

$$\alpha = u \pm i\left(\frac{3i}{8A}\right)^{1/2}, \qquad \beta = v \pm \left(\frac{3i}{8B}\right)^{1/2}. \tag{24}$$

This immediately yields the following proposition.

THEOREM 5. Property III. *A three-parameter family of curves is an Ω family of curves if it possesses the Properties* I, II, *and the Property* III *described as follows. To any point P of the plane, there is associated by Property* II *an equilateral hyperbola H. The four foci of H are related to the point P by a direct conformal transformation.*

Thus an Ω family of curves has been completely characterized geometrically by the Properties I, II, and III. In the next section, we shall give an alternate characterization by the means of the focal lemniscates.

5. **Alternate characterization by focal lemniscates.** Next we shall discuss two Properties, II_1 and III_1, which are respectively equivalent to Properties II and III. First let us note that a *minimal quadrilateral* R is any quadrilateral whose sides are minimal lines. The *center* of R is the intersection of its diagonals. By (19) and (20), we discover the following result.

THEOREM 6. Property II_1. *A three-parameter family of curves with the Property* I *will possess the property that the* ∞^1 *focal lemniscates associated with any point* P *of the plane all pass through the vertices of a minimal quadrilateral* R *with center at* P *if and only if its differential equation is of the form* (22).

The four vertices of the minimal quadrilateral of our family of curves (22) are

$$\alpha = u \pm \left(\frac{3}{8A}\right)^{1/2}, \qquad \beta = v \pm \left(\frac{3}{8B}\right)^{1/2}. \tag{25}$$

From this, we find

THEOREM 7. Property III_1. *A three-parameter family of curves is an* Ω *family of curves if it possesses the Properties* I, II_1 *and the Property* III_1 *described as follows. To any point* P *of the plane, there is associated a minimal quadrilateral* R. *The four vertices of* R *are related to the point* P *by a direct conformal transformation.*

In the above work, we have completely characterized an Ω family of curves by means of the focal lemniscates. That is, we have given *two* equivalent sets of geometric characterizations by means of focal lemniscates. In the next section, we shall characterize any such family by means of directorial equilateral hyperbolas.

6. **Characterization by directorial curves.** In this section, we wish to consider three-parameter families of curves such that the corresponding directorial curve of any lineal element $E(u, v, p)$ of the plane is an equilateral hyperbola H with the point of E as the center of H and the line of E as one of the asymptotes of H. The equation of H must be of the form

$$(\beta - v)^2 - p^2(\alpha - u)^2 = 3p^3/g^2, \tag{26}$$

where g is an arbitrary function of (u, v, p) only. Substituting (15) into (26), we find

$$\frac{-p'f_{p'} + f + 3p'^2/2p}{(-p'f_{p'} + 2f)^2} = \frac{1}{2g^2}. \tag{27}$$

This is a partial differential equation in the unknown function f whose form is to be determined.

Upon making the transformation

$$f = \frac{3p'^2}{2p} + \tfrac{1}{2}(g^2 - F^2), \tag{28}$$

where F is our new unknown function, the differential equation (27) becomes

$$F^2[p'^2F_{p'}^2 - 2p'FF_{p'} + F^2 - g^2] = 0. \tag{29}$$

The solution of this yields

$$F = 0, \quad \text{or} \quad F = \mp g + p'h, \tag{30}$$

where g and h are functions of (u, v, p) only. This shows that our differential equation (13) must be either one of the two types

$$\begin{aligned} &\text{A:} \quad p'' = \frac{3p'^2}{2p} + g^2/2; \\ &\text{B:} \quad p'' = \frac{3p'^2}{2p} \pm ghp' - \frac{h^2}{2}p'^2. \end{aligned} \tag{31}$$

The second type B is any three-parameter family of curves whose differential equation is of the form

$$p'' = Gp' + Hp'^2, \tag{32}$$

where G and H are arbitrary functions of (u, v, p) only. Any such family may be characterized by the property that the focal curve at any element E is a circle passing through the point of E. This type has been considered extensively by Kasner in connection with his geometry of dynamical trajectories (first of Kasner's five dynamical properties(2)).

Upon substituting the second of equations (31) into (14) and (15), we discover that the directrices of the ∞^1 osculating parabolas form a pencil of straight lines with the vertex

$$\alpha = u \pm \frac{3 - h^2p}{2gh}, \qquad \beta = v \pm \frac{(3 + h^2p)p}{2gh}. \tag{33}$$

Thus for this second type B our equilateral hyperbola is degenerate. As a matter of fact, it may be proved(2) that *any three-parameter family of curves for which the* ∞^1 *directrices of the osculating parabolas at any element of the plane form a pencil of straight lines is given by a differential equation of the form* (32).

The first type (A) of (31) is any three-parameter family of curves whose differential equation is of the form (19). Any such family is characterized by

the Property I. Upon substituting the first of equations (31) into (14) and (15), we find that our directorial curve is a nondegenerate equilateral hyperbola as given by (26). By comparing the second of equations (31) with (19), we find $\lambda = pg^2$. Our equilateral hyperbola (26) then becomes

$$(\beta - v)^2 - p^2(\alpha - u)^2 = 3p^4/\lambda. \tag{34}$$

THEOREM 8. Property I′. *A three-parameter family of curves for which the directorial curve at any element E is a nondegenerate equilateral hyperbola H with the point of E as the center of H and the line of E as one of the asymptotes of H is given by a differential equation of the form* (19). *Thus Property* I′ *is equivalent to Property* I.

From (34), we find that if (α, β) denotes any focus of our equilateral hyperbola H, then

$$\begin{gathered} (\alpha - u)^2 = -3p^2/\lambda, \qquad (\beta - v)^2 = 3p^4/\lambda, \\ A(\alpha - u)^2 + B(\beta - v)^2 = -3p^2(A - Bp^2)/\lambda, \end{gathered} \tag{35}$$

where A and B depend only on (u, v). From the above equations, we deduce the following result.

THEOREM 9. Property II′. *A three-parameter family of curves with the Property* I′ *will possess the property that the curve of foci of the directorial equilateral hyperbolas associated with any point P of the plane is an equilateral hyperbola H′ with center at P if and only if its differential equation is of the form* (22). *Thus Property* II′ *is equivalent to Property* II.

The equilateral hyperbolas of Theorem 9 are

$$A(\alpha - u)^2 + B(\beta - v) = -\tfrac{3}{2}. \tag{36}$$

Next if (α, β) denotes the center of any one of the four pairs of orthogonal circles which define any one of the above lemniscates, then

$$\alpha = u \pm i\left(\frac{3}{2A}\right)^{1/2}, \qquad \beta = v \pm i\left(\frac{3}{2B}\right)^{1/2}. \tag{37}$$

THEOREM 10. Property III′. *A three-parameter family of curves is an Ω family of curves if it possesses the Properties* I′, II′, *and the Property* III′ *described as follows. To any point P of the plane, there is associated by Property* II′ *an equilateral hyperbola H′. The four foci of H′ are related to the point P by a direct conformal transformation. Thus Property* III′ *is equivalent to Property* III.

In this section, we have completely characterized an Ω family of curves with a third set of geometric properties. This set depends essentially on the directorial equilateral hyperbolas. In the next section, we shall give some additional geometric properties of such a family.

7. **Additional geometric properties of an Ω family.** Upon replacing λ by the value $2p^2(A-Bp^2)$ in (34), we find that the envelope of the ∞^1 equilateral hyperbolas (34) is

$$[A(\alpha-u)^2+B(\beta-v)^2+\tfrac{3}{4}]^2=4AB(\alpha-u)^2(\beta-v)^2. \tag{38}$$

From this, we derive the following result.

THEOREM 11. *An Ω family of curves possesses the following additional property. The envelope of the ∞^1 directorial equilateral hyperbolas of the lineal elements through a fixed point P consists of the four sides of a rectangle R whose diagonals are the minimal lines through P. Moreover, the vertices of this rectangle R are connected to P by a direct conformal transformation.*

Next we note that the two sets of geometric characterizations (I, II, III) and (I′, II′, III′) give rise to a reciprocity relation. Let us consider as reciprocal elements: the focus and directrix of our osculating parabola; a lemniscate L and an equilateral hyperbola H with the node of L and the center of H coincident; the tangent lines of the node of L and the asymptotes of H. Of course, the centers of the orthogonal pairs of circles defining L and the foci of H both describe equilateral hyperbolas, which are equivalent by a similitude. Then we observe that the two sets of geometric characterizations (I, II, III) and (I′, II′, III′) are reciprocal. This reciprocity is brought to light more by the following two results.

THEOREM 12. *The focal lemniscate L (20) of Property* I *and the directorial equilateral hyperbola H (34) of Property* I′ *are equivalent under the Moebius inversion with respect to the circle with center at the point P and with radius $R^2=3p^3/2\lambda$.*

THEOREM 13. *The equilateral hyperbolas H (23) of Property* II *and H' (36) of Property* II′ *are equivalent under the similitude*

$$\alpha'-u=\pm 2i^{3/2}(\alpha-u), \quad \beta'-v=\pm 2i^{1/2}(\beta-v). \tag{39}$$

This similitude can be factored into the product of the rotation about the point P through the angle $\pi/4+k\pi/2$ by the magnification through P of ratio ± 2 or $\pm 2i$ according as k is odd or even.

In the next section, we shall give a fourth and final geometric characterization of an Ω family of curves by means of the rate of variation of curvature with respect to the arc length.

8. **Characterization by the rate of variation of curvature.** The curvature κ and the rate $d\kappa/ds$ of variation of the curvature with respect to the arc length s are given in minimal coordinates by

$$2i\kappa=\frac{p'}{p^{3/2}}, \quad 2i\frac{d\kappa}{ds}=\frac{2pp''-3p'^2}{2p^3}. \tag{40}$$

From these, we derive the following.

THEOREM 14. Property I''. *A three-parameter family of curves for which the rate* $d\kappa/ds = \kappa'$ *of variation of the curvature per unit length of arc is the same for the* ∞^1 *curves of the family at any lineal element E must have its differential equation of the form* (19). *Thus the Properties* I, I', I'' *are all equivalent.*

Let a three-parameter family of curves possess the Property I''. Then to each lineal element of the plane, there is associated a definite rate of variation of curvature. That is, $\kappa' = \kappa'(E)$.

THEOREM 15. Property II''. *Consider any three-parameter family of curves with the Property* I''. *Let there exist an orthogonal net of curves N with the following property. At any point P of the plane, construct the two orthogonal lineal elements* E_j $(j = 1, 2)$ *which belong to N. The rate* $\kappa'(E)$ *of variation of curvature of any element E through P is proportional to the cosine of twice the angle between E and* E_j. *Any three-parameter family with this additional property must be given by a differential equation of the form* (22).

The Property II'' of the preceding theorem is an analogue of the Meusnier theorem for curves on a surface in three-space. This property may be written in the form

$$\kappa' = \kappa_j' \cos 2\theta, \tag{41}$$

where θ is the angle between the two lineal elements E and E_j at the point P and κ' and κ_j' are the rates of variation of curvature at E and E_j respectively.

The differential equation of the net N and the value of κ_j' are respectively

$$p_j = \pm\, i\left(\frac{A}{B}\right)^{1/2}, \qquad \kappa_j' = \mp\, (AB)^{1/2}. \tag{42}$$

By this, we may state simply

THEOREM 16. Property III''. *A three-parameter family of curves is an* Ω *family if and only if it possesses the Properties* I', II', *and the Property* III' *described as follows. The function* $\kappa_j' p_j$ *depends on u only and the function* κ_j'/p_j *depends on v only.*

This completes our list of geometric characterizations of an Ω family. We have completely listed four such sets of characterizations, namely: Properties (I, II, III); (I, II_1, III_1); (I', II', III'); and (I'', II'', III'').

9. **The hyperosculated isothermal net of an Ω family.** From Theorem 16, we deduce that the orthogonal net N of Theorem 15 is an isothermal net. But this property is not sufficient to characterize an Ω family of curves. Next we make the following observation.

THEOREM 17. *All the curves of an Ω family which are hyperosculated by their osculating circles form an isothermal net. This net cuts the isothermal net N of Theorems* 15 *and* 16 *at an angle of* $\pi/4$ *radians* (45°).

This last theorem is of significant interest for the following reasons. Let us define a conformal transformation, not of the Moebius group, which preserves the maximum number of circles to be a *conformal near-Moebius transformation*(3). Under any such transformation Σ, it results that the only possible circles that are preserved must form the hyperosculated isothermal net of the Ω family into which the ∞^3 circles are converted by Σ. It is known that the only isothermal circular nets are two orthogonal pencils of circles. The conformal near-Moebius transformations are deducible from this result. These are of the form $\Sigma = M_2TM_1$ where M_1 and M_2 are Moebius transformations and T is one of the *three* transformations

$$U = u^n, \quad V = v^n; \qquad U = e^u, \quad V = e^v; \qquad U = \log u, \quad V = \log v. \tag{43}$$

As a corollary of this work, we may state

THEOREM 18. *The only Ω family of curves (not circles) of the (u, v) plane which contain circles as the hyperosculated isothermal net are*

$$\begin{aligned} aU^nV^n + bU^n + cV^n + d &= 0, \\ ae^{U+V} + be^U + ce^V + d &= 0, \\ a \log U \log V + b \log U + c \log V + d &= 0, \end{aligned} \tag{44}$$

where (U, V) are Moebius functions of (u, v).

In the present paper, we have completely characterized the Ω families of curves. One analytic and *four* geometric characterizations, each consisting of three independent properties, have been given. In later work, we shall study certain subfamilies(4) of any Ω family and then develop the geometry of these families with respect to Schwarzian reflection (conformal symmetry)(5). We state in conclusion

THEOREM 19. *Every horn angle contained in an Ω family is conformally equivalent to a circular horn angle, and therefore its measure is $M = \infty$.*

This property belongs to a much larger category of families. The conformal measure of a horn angle is defined as

(3) Kasner and De Cicco, *The conformal near-Moebius transformations*, Bulletin of the American Mathematical Society, vol. 46 (1940), pp. 784–793.

(4) This is related to the theory of *natural families*. See papers by Kasner (1909), Lipke (1912), Douglas (1924), Fialkow (1939), and Struik's recent treatise on differential geometry (1938) where other references can be found.

(5) Kasner, Annals of Mathematics, (2), vol. 38 (1937), pp. 873–877.

$$M_{12} = (\gamma_2 - \gamma_1)/\left(\frac{d\tau_2}{ds} - \frac{d\gamma_1}{ds}\right)$$

where γ denotes curvature. If we demand that this shall be constant, a very extensive type of triple family is obtained.

COLUMBIA UNIVERSITY,
NEW YORK, N. Y.
ILLINOIS INSTITUTE OF TECHNOLOGY,
CHICAGO, ILL.

STRUCTURE OF ABELIAN QUASI-GROUPS

BY

D. C. MURDOCH

INTRODUCTION

One of the most noticeable features of non-associative group-like systems is that the lack of associativity removes nearly all the power from the commutative law. For example, although many properties of groups are retained in systems which satisfy certain generalized associative laws, the addition of the commutative law does not usually reduce these systems to anything analogous to abelian groups. The properties which one usually associates with an abelian group, and which one would wish to retain in any non-associative generalization of this concept, are

(a) Indices may be distributed: $(ab)^r = a^r b^r$.

(b) All subgroups are normal.

(c) The subgroups form a Dedekind structure.

A fourth property which naturally comes to mind is that every abelian group is a direct product of cyclic groups. This, however, does not lend itself to generalization since the cyclic group itself loses all its simplicity when the associative law is relaxed. It will therefore play no part in the considerations of this paper.

A definition of an *abelian quasi-group* which retains the above three properties has previously been given(1). It is a system closed under multiplication, which satisfies the quotient axiom and the generalized associative-commutative law $(ab)(cd) = (ac)(bd)$. It is the purpose of this paper to give a complete account of the structure of these systems. The problem of constructing all abelian quasi-groups is solved in the sense that it is reduced to a group-theoretical problem. It is first shown, by consideration of the problem of extension, that every abelian quasi-group is a direct product of a self-unit quasi-group (one in which every element is a right unit) and one which contains an idempotent element. This latter type can always be constructed by performing certain transformations on the Cayley square of an abelian group, while the self-unit quasi-groups result from two applications of the same process. The results appear to indicate that the classic problems of extension, automorphisms, etc., although more cumbersome to handle, are not essentially more difficult than for abelian groups.

Presented to the Society in part October 28, 1939, under the title *Further theory of quasi-groups*, and in part April 26, 1940, under the title *A characterization of abelian quasi-groups*; received by the editors April 15, 1940.

(1) D. C. Murdoch, *Quasi-groups which satisfy certain generalized associative laws*, American Journal of Mathematics, vol. 61 (1939), pp. 509–522.

The results of §II, on extensions by a factor group with unique right unit, are not essential to the main arguments of the paper. Attention is drawn to the three theorems which give the key to the construction of all abelian quasi-groups by labelling them the first, the second, and the third structure theorems.

I. Abelian quasi-groups and their fundamental properties

1. **Abelian quasi-groups.** We shall understand by an *abelian quasi-group* $\mathfrak{G}$ a system of elements which satisfies the following three postulates:

(A) For any ordered pair of elements a, b of $\mathfrak{G}$ there is a unique product ab which is an element of $\mathfrak{G}$.

(B) If a, b are any two elements of $\mathfrak{G}$, there exist in $\mathfrak{G}$ unique elements x and y such that $ax=b$, $ya=b$.

(C) If a, b, c, d are any four elements of $\mathfrak{G}$, then

$$(ab)(cd) = (ac)(bd). \tag{1}$$

With the exception of the last section, this paper is concerned exclusively with abelian quasi-groups, and we shall therefore use the term quasi-group, except in that section, to mean abelian quasi-group. We shall also confine our attention to finite quasi-groups. With this restriction any subset of $\mathfrak{G}$ which is closed under multiplication will form a subquasi-group. Finally, we shall, where convenient, use the terms subgroup and factor group to mean subquasi-group and factor quasi-group without implying that the systems in question are associative. This will lead to no confusion since we shall have no occasion to consider subquasi-groups or factor quasi-groups which are groups in the ordinary sense.

For convenience of reference we shall list some of the chief properties of abelian quasi-groups previously obtained(2).

1. *If a and b are two elements of a quasi-group $\mathfrak{G}$ and $\phi(a)$, $\psi(a)$ denote two "powers" of a, then*

$$\phi(a)\phi(b) = \phi(ab), \qquad \psi[\phi(a)] = \phi[\psi(a)].$$

This generalizes the index law $(ab)^r=a^rb^r$ of abelian groups. It may be stated more generally as follows:

1 (a). *If $a_1, \cdots, a_n$ and $b_1, \cdots, b_n$ are two sets of elements of $\mathfrak{G}$ and if $\phi(a_1, \cdots, a_n)$ denotes some product formed from the elements $a_1, \cdots, a_n$, then*

$$\phi(a_1, \cdots, a_n)\phi(b_1, \cdots, b_n) = \phi(a_1b_1, \cdots, a_nb_n),$$
$$\psi[\phi(a_1, \cdots, a_n)] = \phi[\psi(a_1), \cdots, \psi(a_n)].$$

This may be proved by induction on the total number of elements multiplied together to form $\phi(a_1, \cdots, a_n)$.

---

(2) For proofs, see D. C. Murdoch, op. cit.

2. *Every element a of $\mathfrak{G}$ has a right unit e_a. All right units of $\mathfrak{G}$ form a subgroup $\mathfrak{R}$ and $a \to e_a$ is a homomorphic mapping of $\mathfrak{G}$ on $\mathfrak{R}$.*

We shall refer to $\mathfrak{R}$ as the right unit group of $\mathfrak{G}$.

3. *If $\mathfrak{H}$ is any subgroup of $\mathfrak{G}$, two cosets $a\mathfrak{H}$ and $b\mathfrak{H}$ are either identical or have no elements in common.*

This follows from the necessary and sufficient conditions for the existence of coset expansions as given by Hausmann and Ore([3]), or it may be proved independently as follows:

If $a\mathfrak{H}$ and $b\mathfrak{H}$ contain common elements, there exist elements h_1 and h_2 of $\mathfrak{H}$ such that $ah_1 = bh_2$. We then have

$$a\mathfrak{H} = (ae_a)(h_1\mathfrak{H}) = (ah_1)(e_a\mathfrak{H}) = (bh_2)(e_a\mathfrak{H}) = (be_a)\mathfrak{H} = b_1\mathfrak{H}.$$

Hence there is an element h_3 of $\mathfrak{H}$ such that $ah_1 = (be_a)h_3$ and by repetition of the above argument, $a\mathfrak{H} = b_n\mathfrak{H}$ where b_n is defined by $b_n = b_{n-1}e_a$ and $b_1 = be_a$. But since $\mathfrak{G}$ is finite there exists an n such that $b_n = b$, and therefore $a\mathfrak{H} = b\mathfrak{H}$.

4. *Every element of a coset $a\mathfrak{H}$ defines the same coset $(ah)\mathfrak{H}$ which is independent of h, and is equal to $a\mathfrak{H}$ if and only if $\mathfrak{H}$ contains e_a. Every subgroup $\mathfrak{H}$ of $\mathfrak{G}$ is normal in the sense that its cosets form a factor group $\mathfrak{G}/\mathfrak{H}$ homomorphic to $\mathfrak{G}$.*

It follows from (1) that left (or right) multiplication by an idempotent element e is an automorphism of $\mathfrak{G}$. The properties of the factor group $\mathfrak{G}/\mathfrak{H}$ may therefore be stated as follows:

5. *If $\mathfrak{R}$ is the right unit group, and $\mathfrak{H}$ any subgroup of $\mathfrak{G}$, the factor group $\mathfrak{G}/\mathfrak{H}$ has a unique right unit if and only if $\mathfrak{H} \supseteq \mathfrak{R}$. The mapping of every coset $a\mathfrak{H}$ on the coset defined by any element of $a\mathfrak{H}$ is an automorphism of $\mathfrak{G}/\mathfrak{H}$ equivalent to right multiplication by the idempotent element $\mathfrak{H}$.*

Finally, it is known that the subgroups of $\mathfrak{G}$ which contain any given subgroup form a Dedekind structure. Hence principal chains between $\mathfrak{G}$ and any fixed minimal subgroup all have the same length. That principal chains terminating in different minimal subgroups need not have the same length is illustrated by the example in §8.

2. **Cyclic quasi-groups.** A cyclic quasi-group is one which is generated by a single element. In general not much can be said about such systems, since the usual machinery for dealing with powers breaks down owing to the lack of the associative law. We shall give two results, however, which hold in the abelian case.

THEOREM 1. *The automorphisms of a cyclic abelian quasi-group form an abelian group.*

Proof. Let $\mathfrak{G}$ be an abelian quasi-group generated by an element a and

([3]) B. A. Hausmann and O. Ore, *Theory of quasi-groups*, American Journal of Mathematics, vol. 59 (1937), pp. 983–1004. See also D. C. Murdoch, op. cit., p. 512.

let S and T be any two automorphisms of $\mathfrak{G}$. Then if $a^S=\phi_1(a)$ and $a^T=\phi_2(a)$, we have

$$a^{ST} = \phi_1[\phi_2(a)], \qquad a^{TS} = \phi_2[\phi_1(a)],$$

whence $ST=TS$ by 1, §1.

It follows from this theorem that a cyclic quasi-group can contain at most one idempotent element, for it is easily shown that the automorphisms of $\mathfrak{G}$ defined by multiplication by two distinct idempotent elements do not commute. More generally, we can prove:

THEOREM 2. *A cyclic abelian quasi-group contains only one minimal subgroup.*

Proof. Let $\mathfrak{N}_1$ and $\mathfrak{N}_2$ be two minimal subgroups of $\mathfrak{G}$ and let $\phi(a)$ be any element of $\mathfrak{N}_1$ where a is a generating element of $\mathfrak{G}$. Now if $\psi(a)$ is any element of $\mathfrak{N}_2$ then $\phi[\psi(a)]=\psi[\phi(a)]$ is an element of both $\mathfrak{N}_1$ and $\mathfrak{N}_2$. Hence $\mathfrak{N}_1$ and $\mathfrak{N}_2$ have a non-void crosscut and since both are minimal they must be identical.

It follows from Theorem 2 that the subgroups of a cyclic quasi-group always form a Dedekind structure.

3. **Direct products.** Given two quasi-groups $\mathfrak{G}$ and $\mathfrak{H}$ it is always possible to construct a third quasi-group $\mathfrak{G}\times\mathfrak{H}$ which we shall call their *direct product.* The direct product is defined as the quasi-group consisting of all element pairs (g, h), where g belongs to $\mathfrak{G}$ and h to $\mathfrak{H}$, in which multiplication is defined by

$$(g_1, h_1)\cdot(g_2, h_2) = (g_1g_2, h_1h_2).$$

The multiplication so defined satisfies (1) and the ordinary laws of direct products hold with one important exception. This is that the component quasi-groups $\mathfrak{G}$ and $\mathfrak{H}$ are not necessarily subgroups of $\mathfrak{G}\times\mathfrak{H}$. If $\mathfrak{H}$ contains an indempotent element, then $\mathfrak{G}$ will be a subgroup of $\mathfrak{G}\times\mathfrak{H}$ but otherwise this will not in general be the case. It should also be noted that the right unit group of $\mathfrak{G}\times\mathfrak{H}$ is the direct product of the right unit groups of $\mathfrak{G}$ and $\mathfrak{H}$.

4. **The right unit series.** Let $\mathfrak{N}_1$ be the right unit group of $\mathfrak{G}$, $\mathfrak{N}_2$ the right unit group of $\mathfrak{N}_1$, $\mathfrak{N}_3$ that of $\mathfrak{N}_2$, and so on. Since $\mathfrak{G}$ is assumed finite, we must finally reach a subgroup $\mathfrak{N}_t$ which is its own right unit group. Such a quasi-group, in which every element is a right unit, shall be called *self-unit*. This series of subgroups,

$$\mathfrak{G} \supset \mathfrak{N}_1 \supset \mathfrak{N}_2 \supset \cdots \supset \mathfrak{N}_t$$

shall be called the *right unit series* of $\mathfrak{G}$, and the number t of distinct successive right unit subgroups in the series, the *length* of $\mathfrak{G}$.

We shall use the letter U throughout to denote the mapping of each element of a quasi-group on its right unit. The right unit series can then be written

$$\mathfrak{G} \supset \mathfrak{G}^{U} \supset \mathfrak{G}^{U^2} \supset \cdots \supset \mathfrak{G}^{U^t}.$$

It should also be noted that the final term $\mathfrak{R}_t = \mathfrak{G}^{U^t}$ of this series is the maximal self-unit subgroup of $\mathfrak{G}$ in the sense that every self-unit subgroup is contained in it. For if $\mathfrak{R}$ is self-unit we have

$$\mathfrak{R} = \mathfrak{R}^{U^t} \subset \mathfrak{R}_t.$$

The factor group $\mathfrak{G}/\mathfrak{R}_t$ will also have length t and its maximal self-unit subgroup will be a single idempotent element.

Let $\mathfrak{F}_i$ be the factor group $\mathfrak{R}_{i-1}/\mathfrak{R}_i$, $i=1, 2, \cdots, t$, where $\mathfrak{R}_0 = \mathfrak{G}$. Then each $\mathfrak{F}_i$ has a unique right unit. Since U is a homomorphism which maps each $\mathfrak{R}_{i-1}$ on $\mathfrak{R}_i$ we have for $i=1, 2, \cdots, t-1$,

$$\frac{\mathfrak{R}_{i-1}}{\mathfrak{R}_i'} \simeq \frac{\mathfrak{R}_i}{\mathfrak{R}_{i+1}} \simeq \mathfrak{F}_{i+1},$$

where $\mathfrak{R}_i'$ is the subgroup of $\mathfrak{R}_{i-1}$ consisting of all elements of $\mathfrak{R}_{i-1}$ which have right units in $\mathfrak{R}_{i+1}$. Evidently $\mathfrak{R}_i' \supseteq \mathfrak{R}_i$ and therefore $\mathfrak{F}_i$ contains a subgroup $\mathfrak{M}_i \simeq \mathfrak{R}_i'/\mathfrak{R}_i$ such that

$$\frac{\mathfrak{F}_i}{\mathfrak{M}_i} \simeq \mathfrak{F}_{i+1}.$$

These results may be summed up as follows:

THEOREM 3. *Let $\mathfrak{G} = \mathfrak{R}_0 \supset \mathfrak{R}_1 \supset \cdots \supset \mathfrak{R}_t$ be the right unit series of $\mathfrak{G}$ and let $\mathfrak{F}_i$ be the factor group $\mathfrak{R}_{i-1}/\mathfrak{R}_i$, $i=1, 2, \cdots, t$. Then each $\mathfrak{F}_i$ has a unique right unit, and is an extension of some quasi-group $\mathfrak{M}_i$ by the factor group $\mathfrak{F}_{i+1}$.*

COROLLARY. *If g is the order of $\mathfrak{G}$ and m_i the order of $\mathfrak{M}_i$, f that of $\mathfrak{F}_t$, and r that of $\mathfrak{R}_t$, then*

$$g = m_1 m_2^2 m_3^3 \cdots m_{t-1}^{t-1} \cdot f^t \cdot r.$$

Any or all of the m_i may be unity, but a quasi-group of length t must be divisible by the tth power of an integer, namely f^t. If each m_i is unity, then all the factor groups $\mathfrak{F}_i$ are isomorphic.

From the preceding it follows that any quasi-group of length t can be obtained from its maximal self-unit subgroup either by a series of t extensions by factor groups having unique right unit, or by a single extension by a factor group of length t whose maximal self-unit subgroup is a single element. We shall therefore turn now to a study of the process of extension. We shall first treat the problem of the extension of any quasi-group $\mathfrak{H}$ by a factor group $\mathfrak{F}$ with unique right unit. The same problem when $\mathfrak{F}$ does not have a unique right unit becomes very complicated. One special case, however, can be completely solved, namely the case in which $\mathfrak{H}$ is self-unit and the maximal self-

unit subgroup of $\mathfrak{F}$ consists of a single element. This case is sufficient for our purposes and will be treated in §III.

II. Extensions by a factor group with unique right unit

5. Extensions which preserve the right unit group. Let $\mathfrak{H}$ be any given quasi-group with right unit group $\mathfrak{R}$ and let $\mathfrak{F}$ be a quasi-group with unique right unit e. It is required to construct an abelian quasi-group $\mathfrak{G}$ containing $\mathfrak{H}$ as subgroup and such that $\mathfrak{G}/\mathfrak{H}\simeq\mathfrak{F}$. We shall consider only the case in which the extended group $\mathfrak{G}$ has the same right unit group as $\mathfrak{H}$, namely $\mathfrak{R}$. It is clear from §3 that one such extension $\mathfrak{G}$ always exists, namely the direct product $\mathfrak{H}\times\mathfrak{F}$. We shall use Latin capitals for elements of $\mathfrak{H}$ and small Greek letters for elements of $\mathfrak{F}$ other than the right unit e.

If $\mathfrak{G}$ is an extension of $\mathfrak{H}$ by $\mathfrak{F}$, having right unit group $\mathfrak{R}$, then $\mathfrak{G}$ splits into cosets modulo $\mathfrak{H}$ of the form

$$\mathfrak{G} = g_e\mathfrak{H} + g_\sigma\mathfrak{H} + g_\tau\mathfrak{H} + \cdots .$$

Here $g_e\in\mathfrak{H}$ and the representatives g_e, g_σ, g_τ, $\cdots$ multiply according to the law

$$g_\sigma g_\tau = g_{\sigma\tau}C_{\sigma,\tau}, \tag{2}$$

where the elements of the factor set $C_{\sigma,\tau}$ belong to $\mathfrak{H}$. Since $\mathfrak{G}$ has right unit group $\mathfrak{R}$, each representative g_σ has a right unit E_σ in $\mathfrak{R}$. Hence every element of $\mathfrak{R}$ is the right unit of at least one element in each coset $g_\sigma\mathfrak{H}$. Thus it is always possible to choose a set of representatives $\{g_\sigma\}$ all of which have the same right unit. Such a set shall be called a *normal* set of representatives.

Taking right units of (2) we have

$$E_\sigma E_\tau = E_{\sigma\tau}C'_{\sigma,\tau},$$

where $C_{\sigma,\tau}$ is the right unit of $C_{\sigma,\tau}$. Hence if $\{g_\sigma\}$ is a normal set of representatives the elements of the factor set must all have the same right unit. If $\mathfrak{H}$ is self-unit, this implies that all $C_{\sigma,\tau}$ are equal, and since any element of $\mathfrak{H}$ may then be chosen as factor set it follows that only one extension is possible in this case, namely the direct product.

If $\mathfrak{H}$ is not self-unit more possibilities occur, but a factor set for any extension can always be chosen so that its elements all have the same right unit. An analysis similar to that of Schreier([4]) for ordinary groups yields the following:

Theorem 4. *Let $\mathfrak{H}$ have right unit group $\mathfrak{R}$ and $\mathfrak{F}$ $(e, \sigma, \tau, \cdots)$ have right*

([4]) O. Schreier, *Über die Erweiterung von Gruppen* I, Monatshefte für Mathematik und Physik, vol. 34 (1926), pp. 165–180, or H. Zassenhaus, *Lehrbuch der Gruppentheorie*, vol. 1, pp. 89–93.

unit e. The necessary and sufficient conditions for the existence of an abelian extension $\mathfrak{G}$ of $\mathfrak{H}$ by $\mathfrak{F}$ having right unit group $\mathfrak{R}$ are that there exist elements $C_{\sigma,\tau}$ in $\mathfrak{H}$ and E_σ in $\mathfrak{R}$ satisfying the following three conditions: $C_{\sigma,e}$ are equal for all σ,

$$E_\sigma E_\tau = E_{\sigma\tau}C'_{\sigma,\tau}, \tag{3}$$

$$C_{\sigma\tau,\mu\rho}[C_{\sigma,\tau}C_{\mu,\rho}]^{S_{\sigma\tau\cdot\mu\rho}} = C_{\sigma\mu,\tau\rho}[C_{\sigma,\mu}C_{\tau,\rho}]^{S_{\sigma\mu\cdot\tau\rho}}, \tag{4}$$

where the operation S_σ is defined by $E_\sigma X^{S_\sigma}=X$.

Proof. The necessity of these conditions follows from (2), E_σ being the right unit of the coset representative g_σ. Condition (4) arises from imposing the law (1) on the extended quasi-group. We shall sketch the proof of sufficiency insofar as it differs from the corresponding proof in the case of groups.

Let $C_{\sigma,\tau}$ and E_σ be given satisfying the three conditions of the theorem. We then construct the set $\mathfrak{G}$ of all elements $g_\sigma A$ where σ belongs to $\mathfrak{F}$ and A to $\mathfrak{H}$. These $g_\sigma A$ are to be considered simply as undefined symbols. We shall show that $\mathfrak{G}$ supplies the required extension when multiplication is defined by

$$(g_\sigma A)(g_\tau B) = g_{\sigma\tau}[C_{\sigma,\tau}(AB)^{S_{\sigma\tau}}]. \tag{5}$$

To prove this it is necessary to show, (i) that $\mathfrak{G}$ contains a subgroup $\overline{\mathfrak{H}}$ isomorphic to $\mathfrak{H}$; (ii) that the right unit of every element $g_\sigma A$ lies in the right unit group $\overline{\mathfrak{R}}$ of $\overline{\mathfrak{H}}$; (iii) that $\mathfrak{G}$ is abelian; and (iv) that $\mathfrak{G}/\overline{\mathfrak{H}}\simeq\mathfrak{F}$.

First, let $\overline{\mathfrak{H}}$ consist of all elements g_eA. Then $\overline{\mathfrak{H}}\simeq\mathfrak{H}$ under the correspondence $g_eA\leftrightarrow C_{e,e}A$. For we have

$$\begin{aligned}(g_eA)(g_eB) &= g_e[C_{e,e}(AB)^{S_e}] \leftrightarrow C_{e,e}[C_{e,e}(AB)^{S_e}]\\ &= (C_{e,e}C_{e,e})[C'_{e,e}(AB)^{S_e}]\\ &= (C_{e,e}C_{e,e})(AB) = (C_{e,e}A)(C_{e,e}B),\end{aligned}$$

since from (3) $C'_{e,e}=E_e$.

To prove (ii) we shall show that every element $g_\sigma A$ of $\mathfrak{G}$ has a right unit g_eX_A, where X_A is defined by

$$C_{e,e}X_A = C_{\sigma,e}X_A = E_\sigma E_A,$$

in which E_A is the right unit of A. For we have from (5)

$$(g_\sigma A)(g_eX_A) = g_\sigma[C_{\sigma,e}(AX_A)^{S_\sigma}].$$

However, using the notation $E_\sigma^{(l)}$ for the left unit of E_σ, we have

$$\begin{aligned}E_\sigma[C_{\sigma,e}(AX_A)^{S_\sigma}] &= (E_\sigma^{(l)}C_{\sigma,e})[E_\sigma(AX_A)^{S_\sigma}]\\ &= (E_\sigma^{(l)}C_{\sigma,e})(AX_A) = (E_\sigma^{(l)}A)(C_{\sigma,e}X_A)\\ &= (E_\sigma^{(l)}A)(E_\sigma E_A) = E_\sigma A,\end{aligned}$$

and therefore $C_{\sigma,e}(AX_A)^{S_\sigma}=A$ and $(g_\sigma A)(g_eX_A)=g_\sigma A$.

Now $g_\sigma X_A$ as an element of $\overline{\mathfrak{H}}$ corresponds to $C_{\sigma,\sigma}X_A$ in $\mathfrak{H}$. But since $C_{\sigma,\sigma}X_A = E_\sigma E_A$, this element lies in $\mathfrak{R}$ and therefore $g_\sigma X_A$ lies in $\overline{\mathfrak{R}}$, the right unit group of $\overline{\mathfrak{H}}$. Hence every element of $\mathfrak{G}$ has a right unit in $\overline{\mathfrak{R}}$ as required.

It is clear from (5) that $\mathfrak{G}/\overline{\mathfrak{H}} \simeq \mathfrak{F}$, and therefore it remains only to prove that $\mathfrak{G}$ is abelian. Since this proof is straightforward, though cumbersome, we shall not give it in full here. It is only necessary to note that the operation S_σ may be distributed over elements of $\mathfrak{H}$ in the following manner:

$$(AB)^{S_\sigma} = A^{S_\sigma} B^{S'_\sigma}, \tag{6}$$

where S'_σ is defined by $E'_\sigma B^{S'_\sigma} = B$, where E'_σ is the right unit of E_σ. This follows from the law (1).

A problem which naturally suggests itself, but which will not be considered here, is the extension of $\mathfrak{H}$ by $\mathfrak{F}$ in such a way that the right unit group of the extension $\mathfrak{G}$ is not $\mathfrak{R}$ but some subgroup of $\mathfrak{H}$ containing $\mathfrak{R}$. Conditions similar to those of Theorem 4 can be derived if the factor group satisfies certain restrictions. The existence of such an extension, however, remains problematical since the direct product does not satisfy the conditions of the problem.

6. **Equivalent factor sets.** Two factor sets $C_{\sigma,\tau}$, $D_{\sigma,\tau}$ and corresponding unit sets E_σ, $\overline{E}_\sigma$ are said to be *equivalent* if there exists in $\mathfrak{H}$ a set of elements A_σ such that

$$\overline{E}_\sigma = E_\sigma A'_\sigma, \qquad C_{\sigma,\tau}(A_\sigma A_\tau)^{S_{\sigma\tau}} = A_{\sigma\tau} D_{\sigma,\tau}^{S_{\sigma\tau}}, \tag{7}$$

where A'_σ is the right unit of A_σ. This equivalence relation is seen to be reflexive on putting $A_\sigma = E_\sigma$ and applying (3). The symmetry and transitivity of the relation are not immediately obvious from (7), but they necessarily follow from the following:

THEOREM 5. *The necessary and sufficient condition that two factor sets $(E_\sigma, C_{\sigma,\tau})$ and $(\overline{E}_\sigma, D_{\sigma,\tau})$ give rise to isomorphic extensions of $\mathfrak{H}$ in which each coset $g_\sigma\mathfrak{H}$ is mapped on $\bar{g}_\sigma\mathfrak{H}$ is that they are equivalent.*

Proof. 1. If $\{g_\sigma\}$ is a set of representatives with right units E_σ and factor set $C_{\sigma,\tau}$, and if $\{\bar{g}_\sigma\}$ is a second set of representatives, corresponding to the same extension, with right units $\overline{E}_\sigma$ and factor set $D_{\sigma,\tau}$, since $\bar{g}_\sigma$ belongs to the coset $g_\sigma\mathfrak{H}$ we have

$$\bar{g}_\sigma = g_\sigma A_\sigma, \qquad A_\sigma \in \mathfrak{H},$$

and therefore $\overline{E}_\sigma = E_\sigma A'_\sigma$. Moreover,

$$\bar{g}_\sigma \bar{g}_\tau = (g_\sigma A_\sigma)(g_\tau A_\tau) = g_{\sigma\tau}[C_{\sigma,\tau}(A_\sigma A_\tau)^{S_{\sigma\tau}}],$$

$$\bar{g}_\sigma \bar{g}_\tau = \bar{g}_{\sigma\tau} D_{\sigma,\tau} = (g_{\sigma\tau} A_{\sigma\tau}) D_{\sigma,\tau} = g_{\sigma\tau}(A_{\sigma\tau} D_{\sigma,\tau}^{S_{\sigma\tau}}).$$

This gives (7), and the factor sets are therefore equivalent.

2. Conversely, if the factor sets are equivalent let $\mathfrak{G}$ and $\bar{\mathfrak{G}}$ be the corresponding extensions. We shall show that the mapping

$$\bar{g}_\sigma A \to g_\sigma(A_\sigma A^{S_\sigma})$$

is an isomorphism between $\mathfrak{G}$ and $\bar{\mathfrak{G}}$. For if

$$\bar{g}_\tau B \to g_\tau(A_\tau B^{S_\tau}),$$

then the product $(\bar{g}_\sigma A)(\bar{g}_\tau B) = \bar{g}_{\sigma\tau}[D_{\sigma,\tau}(AB)^{\bar{S}_{\sigma\tau}}]$ is mapped on the element

$$g_{\sigma\tau}[A_{\sigma\tau}(D_{\sigma,\tau}(AB)^{\bar{S}_{\sigma\tau}})^{S_{\sigma\tau}}] = g_{\sigma\tau}\{(A_{\sigma\tau}D_\sigma^{S_{\sigma\tau}})[A'_{\sigma\tau}(AB)^{\bar{S}_{\sigma\tau}S'_{\sigma\tau}}]\}. \tag{8}$$

However,

$$\begin{aligned}[g_\sigma(A_\sigma A^{S_\sigma})][g_\tau(A_\tau B^{S_\tau})] &= g_{\sigma\tau}\{C_{\sigma,\tau}[(A_\sigma A_\tau)(A^{S_\sigma}B^{S_\tau})]^{S_{\sigma\tau}}\} \\ &= g_{\sigma\tau}\{[C_{\sigma,\tau}(A_\sigma A_\tau)^{S_{\sigma\tau}}][C'_{\sigma,\tau}(A^{S_\sigma}B^{S_\tau})^{S'_{\sigma\tau}}]\}.\end{aligned}$$

Since, by (3), $C'_{\sigma\tau} = (E_\sigma E_\tau)^{S_{\sigma\tau}}$ this is seen by means of (6) to be equal to

$$g_{\sigma\tau}\{[C_{\sigma,\tau}(A_\sigma A_\tau)^{S_{\sigma\tau}}](AB)^{S_{\sigma\tau}}\}.$$

In view of (7) this is identical with the right member of (8) since

$$X = \bar{E}_\sigma X^{\bar{S}_\sigma} = (E_\sigma A'_\sigma)X^{\bar{S}_\sigma} = E_\sigma[A'_\sigma X^{\bar{S}_\sigma S'_\sigma}].$$

The two extensions are therefore isomorphic.

III. Extension by a factor group without unique right unit

7. **The first structure theorem.** It follows from §4 that every abelian quasi-group $\mathfrak{G}$ is an extension of its maximal self-unit subgroup $\mathfrak{N}$ by a factor group $\mathfrak{F}$ having the following two properties:

(a) $\mathfrak{F}$ contains a unique idempotent element e.

(b) If $\mathfrak{F}$ has length t, then U^t maps every element of $\mathfrak{F}$ on e.

We shall now study extensions of this type and shall prove that $\mathfrak{G}$ must in fact be the direct product $\mathfrak{N} \times \mathfrak{F}$.

Let $\mathfrak{N}$ be any self-unit quasi-group and let $\mathfrak{F}$ be a quasi-group having properties (a) and (b). Let $\mathfrak{G}$ be an extension of $\mathfrak{N}$ by $\mathfrak{F}$ and let $\{g_\sigma\}$ be a set of representatives of the cosets of $\mathfrak{G}$ modulo $\mathfrak{N}$. Thus we have

$$\mathfrak{G} = g_e\mathfrak{N} + g_\sigma\mathfrak{N} + g_\tau\mathfrak{N} + \cdots,$$

where $g_e \in \mathfrak{N}$, and where

$$(g_\sigma\mathfrak{N})(g_\tau\mathfrak{N}) = (g_\sigma g_\tau)\mathfrak{N} = g_{\sigma\tau}\mathfrak{N}. \tag{9}$$

Since $\mathfrak{N}$ is not the whole right unit group of $\mathfrak{G}$ the coset $g_\sigma\mathfrak{N}$ does not in general contain g_σ. Let $g_{\sigma S}\mathfrak{N}$ be the coset which contains g_σ. By 5, §1, $g_\sigma\mathfrak{N} \to g_{\sigma S}\mathfrak{N}$ is an automorphism of $\mathfrak{G}/\mathfrak{N}$, namely the inverse of right multiplication by $\mathfrak{N}$. Since $\mathfrak{G}/\mathfrak{N} \simeq \mathfrak{F}$ it follows that S is an automorphism of $\mathfrak{F}$ and

is defined by $\sigma^S e=\sigma$. Thus S leaves invariant every element of $\mathfrak{F}$ whose right unit is e, and every coset $g_\sigma\mathfrak{N}$ which contains its defining element g_σ. From (9) it follows that $g_\sigma g_\tau$ and $g_{\sigma\tau}$ belong to the same coset, namely, $g_{(\sigma\tau)S}\mathfrak{N}$, and hence we have

$$g_\sigma g_\tau = g_{(\sigma\tau)}{}^S C_{\sigma,\tau}, \tag{10}$$

where the elements of the factor set $C_{\sigma,\tau}$ belong to $\mathfrak{N}$.

Owing to property (b) in $\mathfrak{F}$, it follows that the homomorphism U^t when applied to $\mathfrak{G}$ gives a one-to-one mapping of the elements in each coset $g_\sigma\mathfrak{N}$ onto $\mathfrak{N}$. Hence we can choose for the representative g_σ that element of $g_{\sigma S}\mathfrak{N}$ which is mapped on any element of $\mathfrak{N}$ that we please. It is evident, therefore, that in any extension a set of representatives $\{g_\sigma\}$ can be chosen such that each one is mapped by U^t onto the same element of $\mathfrak{N}$, say R.

Assuming such a choice of $\{g_\sigma\}$ to have been made, and applying U^t to (10), we see that each element of the factor set is also mapped on R by U^t. But since $\mathfrak{N}$ is self-unit, U^t is an automorphism of $\mathfrak{N}$, and it follows that all elements of the factor set are equal to that (unique) element of $\mathfrak{N}$ mapped on R by U^t. It follows that any extension of $\mathfrak{N}$ by $\mathfrak{F}$ can be obtained by a factor set all of whose elements are equal, and since the choice of R is arbitrary, any such factor set gives rise to the same extension. Hence there is only one extension possible and this must be the direct product. We have therefore proved:

THEOREM 6. (First structure theorem.) *Every abelian quasi-group is the direct product of its maximal self-unit subgroup by a quasi-group which contains a single idempotent element.*

This theorem reduces the consideration of abelian quasi-groups to that of two types, (i) quasi-groups which contain an idempotent element, (ii) self-unit quasi-groups. We deal with these in order in the following sections.

IV. CONSTRUCTION OF ABELIAN QUASI-GROUPS([6])

8. Quasi-groups containing an idempotent element. The results of this section are based on the following:

THEOREM 7. *Let $\mathfrak{G}$ be an abelian quasi-group and let S and T be any two automorphisms of $\mathfrak{G}$ such that $ST=TS$. If a second operation $\times$ be defined in $\mathfrak{G}$ by means of the equation*

$$a \times b = a^T b^S, \tag{11}$$

then the elements of $\mathfrak{G}$ form a second abelian quasi-group under the operation $\times$.

([6]) Some of the results of this section were proved in the special case of quasi-groups with a unique right unit by A. Suskewitsch, in his paper, *On a generalization of the associative law*, these Transactions, vol. 31 (1929), pp. 204–214.

The proof is immediate since it is only necessary to verify that $(a\times b)\times(c\times d)=(a\times c)\times(b\times d)$. We shall use the notation $(\mathfrak{G}, T, S)$ to denote the quasi-group in which multiplication is defined by (11). The theorem then states that if the permutations T^{-1} and S^{-1} are performed on the vertical and horizontal title lines respectively of the Cayley square for $\mathfrak{G}$, the resulting square is still that of an abelian quasi-group, namely $(\mathfrak{G}, T, S)$. It is clear too that T and S are automorphisms of $(\mathfrak{G}, T, S)$ as well as of $\mathfrak{G}$. More generally, any automorphism of $\mathfrak{G}$ which commutes with both T and S is also an automorphism of $(\mathfrak{G}, T, S)$. Thus if $\overline{\mathfrak{G}}=(\mathfrak{G}, T, S)$ we may write $\mathfrak{G}=(\overline{\mathfrak{G}}, T^{-1}, S^{-1})$. However both $\mathfrak{G}$ and $\overline{\mathfrak{G}}$ will in general have automorphisms which are not automorphisms of the other.

In particular the above theorem holds when $\mathfrak{G}$ is an abelian group. This suggests the following:

Definition. *If there exist in a quasi-group $\mathfrak{G}$ two commutative automorphisms T^{-1}, S^{-1} such that $\overline{\mathfrak{G}}=(\mathfrak{G}, T^{-1}, S^{-1})$ is an abelian group, then $\mathfrak{G}\simeq(\overline{\mathfrak{G}}, T, S)$ shall be called a group representation of $\mathfrak{G}$ or simply a representation.*

Let $\mathfrak{G}=(\overline{\mathfrak{G}}, T, S)$ where $\overline{\mathfrak{G}}$ is a quasi-group with commutative automorphisms S and T. Let $\mathfrak{H}$ be any subgroup of $\overline{\mathfrak{G}}$. The elements of $\mathfrak{H}$ will also form a subgroup of $\mathfrak{G}$ if and only if $a^Tb^S\in\mathfrak{H}$ for all a and b in $\mathfrak{H}$. A sufficient condition for this is that $\mathfrak{H}$ be invariant under T and S. If $\mathfrak{H}$ contains the right unit group $\overline{\mathfrak{R}}$ of $\overline{\mathfrak{G}}$ this condition is also necessary. For in that case

$$a^{T} e_a^{S} \in \mathfrak{H}$$

for all a in $\mathfrak{H}$. But, since $\overline{\mathfrak{R}}$ is clearly a characteristic subgroup e_a^S belongs to $\overline{\mathfrak{R}}$, and therefore to $\mathfrak{H}$. Hence $a^T\in\mathfrak{H}$ for all a in $\mathfrak{H}$ and similarly a^S also belongs to $\mathfrak{H}$. We have therefore proved:

Theorem 8. *Let $\overline{\mathfrak{G}}$ be a quasi-group with right unit group $\overline{\mathfrak{R}}$ and commutative automorphisms T and S. If $\mathfrak{H}$ is a subgroup of $\overline{\mathfrak{G}}$ containing $\overline{\mathfrak{R}}$, then the necessary and sufficient condition that $\mathfrak{H}$ be also a subgroup of $(\overline{\mathfrak{G}}, T, S)$ is that it be invariant under T and S.*

The following theorem enables us to construct all abelian quasi-groups which contain an idempotent element, from abelian groups.

Theorem 9. (Second structure theorem.) *Every abelian quasi-group $\mathfrak{G}$ which contains an idempotent element has a representation of the form $(\overline{\mathfrak{G}}, T, S)$ where $\overline{\mathfrak{G}}$ is an abelian group*(6).

(6) This theorem exposes a misstatement in the author's previous paper (loc. cit., p. 522), where an example is given with the statement that it cannot be constructed from an abelian group by this method. The example in question can be so constructed from the direct product of two cyclic groups of order 3.

Proof. The proof consists in showing that there exist automorphisms T and S of $\mathfrak{G}$ such that $(\mathfrak{G}, T^{-1}, S^{-1})$ is an abelian group. Let e be an idempotent element of $\mathfrak{G}$ and denote by S and T respectively the automorphisms defined by left and right multiplication by e. We then have

$$a^T = ae, \qquad a^S = ea. \tag{12}$$

Since $e(ae) = (ea)e$, we have $ST = TS$. Moreover,

$$a^T(bc^{S^{-1}}) = (ae)(bc^{S^{-1}}) = (ab)(ec^{S^{-1}}) = (ab)c,$$

and therefore an abelian quasi-group containing an idempotent element satisfies the associative law

$$(ab)c = a^T(bc^{S^{-1}}). \tag{13}$$

We now define the operation $\times$ by means of the equation

$$a \times b = a^{T^{-1}}b^{S^{-1}}.$$

On applying (13) it is easily seen that

$$(a \times b) \times c = a \times (b \times c).$$

Finally, since $a \times e = a^{T^{-1}}e = a$, and $e \times a = ea^{S^{-1}} = a$, it is clear that under the operation $\times$ the elements of $\mathfrak{G}$ form a group with unit element e. Moreover it is an abelian group since by Theorem 7 it is an abelian quasi-group. Denoting this group by $\overline{\mathfrak{G}} = (\mathfrak{G}, T^{-1}, S^{-1})$, we have then $\mathfrak{G} = (\overline{\mathfrak{G}}, T, S)$, which is the required representation of $\mathfrak{G}$.

The converse of the above theorem, namely that every quasi-group of the form $(\mathfrak{G}, T, S)$, where $\mathfrak{G}$ is an abelian group, contains an idempotent element, is certainly true since the unit element of $\mathfrak{G}$ will remain idempotent in $(\mathfrak{G}, T, S)$. However quasi-groups so formed may contain subgroups which have no idempotent element. This is illustrated by the example given below:

		1	2	3	4	5	6	7	8	9
	1	1	2	3	4	5	6	7	8	9
	2	2	3	1	5	6	4	8	9	7
	3	3	1	2	6	4	5	9	7	8
	4	4	5	6	7	8	9	2	3	1
(α)	5	5	6	4	8	9	7	3	1	2
	6	6	4	5	9	7	8	1	2	3
	7	7	8	9	2	3	1	5	6	4
	8	8	9	7	3	1	2	6	4	5
	9	9	7	8	1	2	3	4	5	6

	·	1	2	3	4	5	6	7	8	9
	1	1	3	2	8	7	9	4	6	5
	2	3	2	1	7	9	8	6	5	4
	3	2	1	3	9	8	7	5	4	6
	4	7	9	8	6	5	4	2	1	3
(β)	5	9	8	7	5	4	6	1	3	2
	6	8	7	9	4	6	5	3	2	1
	7	5	4	6	1	3	2	8	7	9
	8	4	6	5	3	2	1	7	9	8
	9	6	5	4	2	1	3	9	8	7

Table (α) is the cyclic group $\mathfrak{G}$ of order 9, while table (β) is the quasi-group

$(\mathfrak{G}, S, S^{-1})$ where S is the automorphism which maps each element of $\mathfrak{G}$ on its square. This quasi-group contains three subgroups of order 3, one having three idempotent elements and the other two none.

If $\bar{\mathfrak{G}}$ is an abelian group, then by Theorem 8 every subgroup of $\bar{\mathfrak{G}}$ which is invariant under T and S is also a subgroup of $(\bar{\mathfrak{G}}, T, S)$. Similarly every subgroup of $(\bar{\mathfrak{G}}, T, S)$ which is invariant under T and S (i.e., every subgroup which contains e, the unit of $\bar{\mathfrak{G}}$) is also a subgroup of $\bar{\mathfrak{G}}$. We therefore have a structure isomorphism between the structure Σ_1 of all subgroups of $(\bar{\mathfrak{G}}, T, S)$ which contain e, and the structure Σ_2 of all subgroups of $\bar{\mathfrak{G}}$ which are invariant under T and S.

The idempotent elements of $\mathfrak{G}$ form a subgroup $\mathfrak{Q}$ in which every element is its own right (and left) unit. If the elements of $\mathfrak{Q}$ are denoted by $e_1, e_2, \cdots$, then corresponding to each e_i we can construct a representation of $\mathfrak{G}$ as in Theorem 9. Thus if $e_i a = a^{S_i}$ and $ae_i = a^{T_i}$, then $(\mathfrak{G}, T_i^{-1}, S_i^{-1})$ is an abelian group with unit element e_i.

THEOREM 10. *The abelian groups $(\mathfrak{G}, T_i^{-1}, S_i^{-1})$ constructed as in Theorem 9 from the different idempotent elements of $\mathfrak{G}$ are all isomorphic.*

Proof. Let e_1 and e_2 be two idempotent elements and T_1, S_1, and T_2, S_2 the corresponding automorphisms of $\mathfrak{G}$. Let $\mathfrak{G}_1 = (\mathfrak{G}, T_1^{-1}, S_1^{-1})$ and $\mathfrak{G}_2 = (\mathfrak{G}, T_2^{-1}, S_2^{-1})$. Any isomorphic mapping of $\mathfrak{G}_1$ on $\mathfrak{G}_2$ must certainly map e_1 on e_2 since these are the unit elements of the two groups. We shall show that such a mapping is furnished by the automorphism $S_1T_2S_1^{-1}$ of $\mathfrak{G}$.

First, $S_1T_2S_1^{-1}$ takes e_1 into e_2 since $e_1^{S_1T_2} = e_2^{S_1} = e_1e_2$. It remains to show therefore that

$$(a^{T_1^{-1}}b^{S_1^{-1}})^{S_1T_2S_1^{-1}} = (a^{S_1T_2S_1^{-1}})^{T_2^{-1}}(b^{S_1T_2S_1^{-1}})^{S_2^{-1}},$$

or, in other words, that in the automorphism group of $\mathfrak{G}$, T_2 is the transform of T_1 by $S_1T_2S_1^{-1}$ and similarly S_2 is the transform of S_1 by $S_1T_2S_2^{-1}$. This follows since for all a

$$\begin{aligned} a^{S_1T_2S_1^{-1}T_2S_1} &= e_1[((e_1a)e_2)^{S_1^{-1}}e_2] = [(e_1a)e_2](e_1e_2) \\ &= [(e_1a)e_1]e_2 = [e_1(ae_1)]e_2 = a^{T_1S_1T_2}. \end{aligned}$$

Hence $S_1T_2S_1^{-1}T_2S_1 = T_1S_1T_2$ or $T_2 = (S_1T_2S_1^{-1})^{-1}T_1(S_1T_2S_1^{-1})$ and similarly $S_2 = (S_1T_2S_1^{-1})^{-1}S_1(S_1T_2S_1^{-1})$.

It is easily shown that if $(\mathfrak{G}, T, S)$ is a group it must be one of the groups defined, in the manner of Theorem 9, by means of an idempotent element of $\mathfrak{G}$. For the unit element e of $(\mathfrak{G}, T, S)$ must be invariant under T and S and therefore is an idempotent element of $\mathfrak{G}$. That T and S are the automorphisms defined by (12) readily follows. Hence from Theorem 10 it follows that a quasi-group can be represented in the form $(\bar{\mathfrak{G}}, T, S)$ by only one abelian group $\bar{\mathfrak{G}}$.

9. **Further transformations of the Cayley square.** We have seen that if $\mathfrak{G}$ contains an idempotent element e and if S and T are defined by (12), then $(\mathfrak{G}, T^{-1}, S^{-1})$ is an abelian group. We shall now show that the restrictions that e be idempotent and S and T automorphisms are not essential. Let g be an arbitrary element of $\mathfrak{G}$ and define the operations S_g and T_g by the equations

$$a^{S_g} = ga, \qquad a^{T_g} = ag.$$

However for typographical reasons we shall drop the subscripts and write $S=S_g$ and $T=T_g$. Using S' and S'' to denote left multiplication by the right unit e_g and the left unit e_g' respectively, and T', T'' to denote right multiplication by the same elements, we have from (1) the following distributive laws:

$$(ab)^S = a^S b^{S'} = a^{S''} b^S, \qquad (ab)^T = a^T b^{T'} = a^{T''} b^T.$$

The same laws hold for the inverse operations S^{-1}, T^{-1}. Thus S and T can be thought of as "pseudo-automorphisms" which become true automorphisms if g is idempotent. We can now generalize Theorem 9 as follows:

THEOREM 11. *If $\mathfrak{G}$ is any abelian quasi-group, and if S and T respectively denote left and right multiplication by a fixed element g, then $(\mathfrak{G}, T^{-1}, S^{-1})$ is an abelian group.*

Proof. If we define the operation $\times$ by

$$a \times b = a^{T^{-1}} b^{S^{-1}},$$

we then have

$$(a \times b) \times (c \times d) = (a^{T^{-1}} b^{S^{-1}})^{T^{-1}} (c^{T^{-1}} d^{S^{-1}})^{S^{-1}} \tag{14}$$
$$= (a^{T^{-2}} b^{S^{-1}T'^{-1}})(c^{T^{-1}S''^{-1}} d^{S^{-2}}).$$

But from (1), $g(ae_g)=(e_g' a)g$ for all a, and therefore $T'S=S''T$, or $S^{-1}T'^{-1}=T^{-1}S''^{-1}$. Thus (14) is symmetric in b and c, and hence $(a\times b)\times(c\times d)=(a\times c)\times(b\times d)$. Moreover $(\mathfrak{G}, T^{-1}, S^{-1})$ has a unit element, namely, gg. For

$$a \times gg = a^{T^{-1}} g = a, \qquad gg \times a = g a^{S^{-1}} = a.$$

Hence $(\mathfrak{G}, T^{-1}, S^{-1})$ is an abelian quasi-group with unit element and therefore an abelian group.

It is easy to show similarly that $(\mathfrak{G}, 1, S_g)$, $(\mathfrak{G}, 1, S_g^{-1})$, $(\mathfrak{G}, T_g, 1)$ and $(\mathfrak{G}, T_g^{-1}, 1)$ are all abelian quasi-groups. In the second of these g is a unique left unit and in the fourth, a unique right unit. These results may be stated in terms of transformations of the Cayley square as follows:

Given the Cayley square of an abelian quasi-group $\mathfrak{G}$,

(i) $(\mathfrak{G}, 1, S_g^{-1})$ is obtained by replacing the horizontal title line by the gth row;

(ii) $(\mathfrak{G}, T_g^{-1}, 1)$ is obtained by replacing the vertical title line by the gth column;

(iii) $(\mathfrak{G}, T_g^{-1}, S_g^{-1})$ is obtained by performing the replacements (i) and (ii) simultaneously;

(iv) $(\mathfrak{G}, 1, S_g)$ is obtained by replacing each element x of the horizontal title line by the element of that line, which occurs above the element x in the gth row;

(v) $(\mathfrak{G}, T_g, 1)$ is obtained by replacing each element x in the vertical title line by the element in that line, which occurs to the left of the element x in the gth column.

10. **Self-unit quasi-groups.** In view of Theorems 6 and 9 the only remaining problem in the characterization of all abelian quasi-groups is to construct those self-unit quasi-groups which do not contain an idempotent element. If $\mathfrak{G}$ is any such quasi-group, the mapping U, of each element on its right unit, is an automorphism of $\mathfrak{G}$ which leaves no element unchanged. If we define a new operation $\times$ in $\mathfrak{G}$ by means of the equation

$$a \times b = ab^U = ae_b,$$

it is clear that in the resulting quasi-group $(\mathfrak{G}, 1, U)$ every element is idempotent. Moreover U is also an automorphism of $(\mathfrak{G}, 1, U)$. We may state this result as follows:

THEOREM 12. (Third structure theorem.) *Every self-unit quasi-group which does not contain an idempotent element has the form $\mathfrak{G} = (\overline{\mathfrak{G}}, 1, U^{-1})$, where $\overline{\mathfrak{G}}$ is a quasi-group in which every element is idempotent, and U an automorphism of $\overline{\mathfrak{G}}$ which changes every element. Moreover U is the automorphism of $\mathfrak{G}$ which maps each element on its right unit.*

Thus every abelian quasi-group has one of the following three forms:

(a) $(\mathfrak{G}, T, S)$, where $\mathfrak{G}$ is an abelian group, and S and T commutative automorphisms of $\mathfrak{G}$;

(b) $(\overline{\mathfrak{G}}, 1, S)$, where $\overline{\mathfrak{G}}$ has the form (a) and S is an automorphism of $\overline{\mathfrak{G}}$;

(c) direct products of types (a) and (b).

The quasi-group $\mathfrak{G} = (\overline{\mathfrak{G}}, 1, U)$, where $\overline{\mathfrak{G}}$ is self-unit and $a^U = e_a$, is again self-unit and contains only idempotent elements. In such a quasi-group left and right multiplication by any element are automorphisms. Let $e_1, e_2, e_3, \cdots$ be the elements of $\mathfrak{G}$, and denote by S_i and T_i the automorphisms defined respectively by left and right multiplication by e_i, and by $S_{i,j}$, $T_{i,j}$ the automorphisms corresponding to the product e_ie_j. The following relations are then easily verified:

$$(15) \qquad S_{i,j} = S_i^{-1}S_jS_i = T_j^{-1}S_iT_j, \qquad T_{i,j} = T_j^{-1}T_iT_j = S_i^{-1}T_jS_i.$$

Thus although $S_iT_i = T_iS_i$, S_i cannot commute with S_j or T_j for $j \neq i$.

If $\mathfrak{A}$ is the automorphism group of $\mathfrak{G}$, then any element of $\mathfrak{A}$ which commutes with all S_i must be the identity. For if $(e_ia)^S = e_ia^S$ we have $e_i^S = e_i$. Now from (15) left multiplication by a product e_ie_j is the transform of S_j by S_i. Applying this, we find that left multiplication by $(e_ie_j)(e_ke_m)$ is equivalent to the automorphism

$$S_i^{-1}S_j^{-1}S_iS_k^{-1}S_mS_kS_i^{-1}S_jS_i.$$

Since by (1) this must be unaltered by the interchange of j and k we find, on equating the two expressions, that for all m

$$S_mS_kS_i^{-1}S_jS_k^{-1}S_iS_j^{-1} = S_kS_i^{-1}S_jS_k^{-1}S_iS_j^{-1}S_m,$$

and therefore

$$S_kS_i^{-1}S_jS_k^{-1}S_iS_j^{-1} = 1, \qquad T_kT_i^{-1}T_jT_k^{-1}T_iT_j^{-1} = 1, \tag{16}$$

the second relation following in a similar manner. Thus although a commutator $S_iS_jS_i^{-1}S_j^{-1}$ cannot be equal to the unit element for $i \neq j$ the "three element commutators" of the form (16) are all equal to the unit.

The following two examples will illustrate the large number of automorphisms enjoyed by quasi-groups of this type:

(a)

	1	2	3
1	1	3	2
2	3	2	1
3	2	1	3

(b)

	1	2	3	4	5
1	1	5	4	3	2
2	3	2	1	5	4
3	5	4	3	2	1
4	2	1	5	4	3
5	4	3	2	1	5

In (a) we have the rather surprising situation that every permutation of the elements is an automorphism. The automorphism group is therefore the full symmetric group. In (b) we have $T_1 = (2\ 3\ 5\ 4)$, $T_2 = (1\ 5\ 3\ 4)$, $T_3 = (1\ 4\ 5\ 2)$, $T_4 = (1\ 3\ 2\ 5)$, $T_5 = (1\ 2\ 4\ 3)$, $S_i = T_i^2$ and $S_1S_2 = (1\ 3\ 5\ 2\ 4)$. These, with their powers, give twenty automorphisms. A quasi-group of this type always contains an automorphism which changes every element. For example, $a^{S_1S_2^{-1}} \neq a$ for all a since $e_1a \neq e_2a$. Hence every such quasi-group gives rise to another, namely $(\mathfrak{G}, 1, S_1S_2^{-1})$, which is self-unit but contains no idempotent element.

V.

11. Some non-abelian quasi-groups. In Theorem 9 it was shown that an abelian quasi-group which contains an idempotent element e satisfies the associative law

$$(ab)c = a^T(bc^{S^{-1}}), \tag{17}$$

where S and T are commutative automorphisms defined by (12). It is natural, therefore, to consider quasi-groups which satisfy (17) but which are not necessarily abelian.

Let $\mathfrak{G}$ be a quasi-group, with commutative automorphisms S and T, which satisfies (17). It follows as in Theorem 9 that the operation $\times$ defined by

$$(18) \qquad a \times b = a^{T^{-1}} b^{S^{-1}}$$

is associative. The elements of $\mathfrak{G}$ therefore form, under this operation, a group[7] which we shall denote by $\overline{\mathfrak{G}}$. Since S and T are also automorphisms of $\overline{\mathfrak{G}}$ the unit element e of $\overline{\mathfrak{G}}$ is invariant under both S and T, and hence by (18), e is an idempotent element of $\mathfrak{G}$ and S and T are defined by (12). Hence every quasi-group which satisfies the associative law (17) has the form $(\overline{\mathfrak{G}}, T^{-1}, S^{-1})$ where $\overline{\mathfrak{G}}$ is a group (not necessarily abelian), and T and S are commutative automorphisms of $\overline{\mathfrak{G}}$.

The transform a_c of a by c can be defined in $\mathfrak{G}$ by the equation

$$c^{T^{-1}} a_c^{S^{-1}} = a^{T^{-1}} c^{S^{-1}}.$$

We then have

$$(ab)_c = a_m b_n$$

where $m = c^{T^{-1}}$ and $n = c^{S^{-1}}$, and it is easily shown that those elements which are transformed into themselves by all elements of $\mathfrak{G}$ form an abelian subquasi-group, the center of $\mathfrak{G}$. These are, of course, exactly the elements of the center of $\overline{\mathfrak{G}}$. A quasi-group satisfying (17) is therefore abelian if and only if

$$a^{T^{-1}} b^{S^{-1}} = b^{T^{-1}} a^{S^{-1}},$$

for all elements a and b of $\mathfrak{G}$.

THEOREM 13. *In a quasi-group satisfying* (17) *there exist coset expansions with respect to any subquasi-group* $\mathfrak{H}$ *which is invariant under* T.

Proof. Let h_1 and h_2 be elements of $\mathfrak{H}$ and let $ah_1 = bh_2$. Then for any element ah of the coset $a\mathfrak{H}$ we have, since $\mathfrak{H}^T = \mathfrak{H}$,

$$ah = a(h_1^{T^{-1}} h_3) = (a^{T^{-1}} h_1^{T^{-1}}) h_3^S = (bh_2)^{T^{-1}} h_3^S = b(h_2^{T^{-1}} h_3) = bh_4.$$

Hence $a\mathfrak{H} \subset b\mathfrak{H}$, and similarly $b\mathfrak{H} \subset a\mathfrak{H}$. Therefore two cosets which have an element in common are identical.

Since T is defined by (12), $\mathfrak{H}$ is invariant under T if and only if it contains the idempotent element e. Thus coset expansions exist for all subquasi-groups of $\mathfrak{G}$ which contain e. A subquasi-group $\mathfrak{H}$ containing e is said to be *normal* if

$$a^{T^{-1}} \mathfrak{H} = \mathfrak{H}^{S^{-1}} a,$$

(7) The associative law together with the left and right quotient axioms imply the existence of a unit element.

for all a in $\mathfrak{G}$. This condition is necessary and sufficient for the existence of a quotient group $\mathfrak{G}/\mathfrak{H}$ homomorphic to $\mathfrak{G}$. It is clear that the normal subquasi-groups of $\mathfrak{G}$ which contain e are simply the normal subgroups of the group $\overline{\mathfrak{G}}=(\mathfrak{G}, T, S)$ and therefore the usual theorems concerning these will carry over to the quasi-group.

UNIVERSITY OF SASKATCHEWAN,
SASKATOON, CANADA.

CONCERNING THE DECOMPOSITION AND AMALGAMATION OF POINTS, UPPER SEMI-CONTINUOUS COLLECTIONS, AND TOPOLOGICAL EXTENSIONS

BY
R. G. LUBBEN

1. **Introduction.** A necessary and sufficient condition that a space H Fréchet have the Borel-Lebesgue covering property is that each monotonic collection of closed point sets in it have a nonvacuous product(1). Thus, if such a space is not perfectly compact, there exists in it a monotonic collection of closed sets, E, with a vacuous product; E is an example of what we call a boundary element. With the help of E we may define a boundary point, $P(E)$; if $P(E)$ is added to the basic space and a suitable topology is introduced, the closures of the set of elements of E contain $P(E)$; and the deficiency of the elements of E, that their product is vacuous, no longer holds for the aggregate of their closures in the extended space. By adding to S an aggregate of boundary points which satisfy suitable conditions we achieve the embedding of S in a perfectly compact Hausdorff space (cf. Theorem 16.1).

In Chapter I we introduce the class of *point elements*; these include the boundary elements. We define the relation of the *intersection* of two point elements; in terms of this we give an ordering which makes the aggregate of all point elements a quasi-partially ordered system. A repetition of this algorithm gives the quasi-partially ordered system M of all collections of point elements. By a process identifying the equivalent elements of M we obtain a partially ordered system whose elements we call *portions* of our basic space S, or *S-portions*; these are our ideal points. A relation of an open set D in S and an ideal point P is expressed by saying that *D is an S-neighborhood of P*; the S-neighborhoods are used in topologizing collections of ideal points. The relation *P is an end of M* between the ideal point P and the closed point set of S, M,

Presented to the Society, in three parts: September 9, 1937, under the title *Perfectly compact Hausdorff spaces in which a normal space may be embedded*; December 28, 1937, under the title *Concerning upper semi-continuous collections and the decomposition of points of normal spaces*; and September 6, 1938, under the title *The extension of homeomorphisms of normal spaces to topologically related spaces*; received by the editors July 18, 1938, and, in revised form, September 9, 1940.

(1) Cf. Moore, (II) (the bibliography is at the end of the paper). Kuratowski and Sierpiński, loc. cit., show that the Borel-Lebesgue property is equivalent to the condition that each infinite point set have a complete limit point. Fréchet, (II), calls this property *perfect compactness*; several years later Alexandroff and Urysohn, (I), call it *bicompactness*.

A collection of sets is said to be *monotonic* provided that if K and L are two of its elements, then either $K \supset L$ or $L \supset K$; Moore, loc. cit.

suggests an analogy between Carathéodory's theory of ends and ours in that both are concerned with methods of approach or accessibility. Our end-theory is used as a technique in developing that of the S-neighborhoods.

We shall now indicate some of the ideas underlying the methods which we discussed in the preceding paragraph. The theory of point elements and that of ends suggest our regarding a point not as a static entity but rather as that of a relation of the point to the remainder of space, that involves methods of approach. Secondly, a point need not necessarily be indivisible. We have a conception according to which our space S consists of a quantity of basic matter; this may be decomposed and put together in various ways, and the same applies for each portion of this matter. In particular, the points of S are subject to such operations. There exists a maximal portion, that of all the matter; there prove to be atomic portions.

These ideas may be described by means of our ordered system of ideal points; let "$A < B$" be the ordering relation of this system. Let P be an ideal point and M be a collection of ideal points; if both $P < M$ and $M < P$, we shall think of P as a *summation* or an *amalgamation* of M, and of M as a breaking up or a *decomposition* of P; we think of M and P as describing different structures of the *same* portion of the basic matter of S. Let X and Y be collections of S-portions and "$y=\alpha(x)$" be a transformation from X to Y such that for $y_1 \in Y$, y_1 is the amalgamation of the elements of $\alpha^{-1}(y_1)$(2); in proceeding from $\alpha^{-1}(y_1)$ to y_1 we are involved in a change of structure, only; and, thus, this transformation may be said to *preserve S-portions* or to *leave S-portions invariant*.

If we were concerned with order relations, only, the system of all our ideal points would afford a satisfactory basis for the amalgamation and decomposition theory. From a topological standpoint, however, we require systems of elements for which the S-portion preserving transformations, $y=\alpha(x)$, which describe the amalgamation and the decomposition processes, satisfy continuity conditions. As illustrations of systems which meet this requirement we refer to a result by Stone(3). Of particular interest is the case where X and Y are perfectly compact Hausdorff spaces; this case leads us to apply a mapping theory by Kolmogoroff, which, for our purposes, finds a summary in Theorem VIII, p. 98, Alexandroff and Hopf, loc. cit. The adaption of this theory to the S-portion preserving transformations leads to the introduction of a special class of ideal points, the *amalgamation points*. It is interesting to note that since in applications we are to use both order and continuity relations, such relations are used in defining these points. Chapter II is devoted mainly to a development of the properties of amalgamation points; such properties have striking analogies to those of closed point sets in perfectly compact Hausdorff spaces. For instance, a decomposition of an amalgamation point into atomic

(2) The symbol, $\alpha^{-1}(y_1)$ means all elements of X, x_1, such that $y_1=\alpha(x_1)$.
(3) Loc. cit., p. 476, Theorem 88; we discuss this in §§16 and 20.

S-portions is a perfectly compact set (cf. Theorem 14.4). In §15 this analogy is shown to be fundamental; here we show that the decomposition M of an amalgamation point P into regular amalgamation points is perfectly compact if and only if it is upper semi-continuous relative to S; we develop an elaboration of the Kolmogoroff-Alexandroff-Hopf theory of the S-portion preserving continuous transformations of such decompositions.

The sections following §15 are devoted to applying these results. In Chapter III we investigate systems of such applications. In §19 we consider the ordered system $\delta(P)$ of all perfectly compact decompositions of P into regular amalgamation points; the ordering "$X<Y$," which means that X may be mapped on Y by a continuous, S-portion preserving mapping $y=\alpha(x)$, makes $\delta(P)$ a complete lattice. If δ is a sublattice of $\delta(P)$, the sum of the elements of δ, L_δ(4), becomes a lattice of amalgamation points. The zero of δ, ω_δ, is the system of atomic elements of L_δ, and may be mapped on each element of δ. Such ω_δ's may be described as universal inverse mapping spaces relative to the elements of δ. Particularly important is the case for which P is the maximal amalgamation point; that is, P is the amalgamation of all our ideal points. Then we let $\delta(S)=\delta(P)$; in §20 we show that the completely regular spaces S are characterized by the property that there exist elements of $\delta(S)$ in which the decomposition of S has proceeded to elements at least as small as the points of S. For such a space we have the interesting subsystem of $\delta(S)$, $\delta=H(S)$, which consists of the images of all immediate extensions of S which are perfectly compact Hausdorff spaces. The zero of $H(S)$, $\lambda(S)$, is the topological image of the space considered by Stone in his Theorem 88 (loc. cit., p. 476). In §17 the space $\lambda(S)$ enters into an interesting characterization of the normality of S; this is given by the condition that each upper semi-continuous decomposition of S into a collection of closed point sets can be extended to a similar decomposition of $\lambda(S)$. Further, only the topological images of $\lambda(S)$ have this property.

In Chapter II we introduce the concept of a semi-completely normal space. For such a space the theory of Chapter III is especially simple; such a space is characterized by the property that the zero of $\delta(P)$, $\omega(P)$, is a set of atomic ideal points. Similarly, the normal space is characterized by the condition that the only boundary points which are elements of an $\omega(P)$ are atomic points. In Chapter II we give characterizations of various types of normal and regular spaces.

Chapter I. An ordered system of ideal points

At the beginning of the preceding section we discussed the methods and the underlying ideas which occur in the development of our theory of point

(4) It may be necessary to add a zero to L_δ to make it a lattice; such is *not* the case for $\delta(P)$, or for δ.

elements and ideal points. It may be mentioned that methods similar to those used in this chapter may be used to extend a partially ordered system K to a complete partially ordered system, granting that K satisfies suitable conditions.

2. **Concerning a topological background.** We refer the reader to treatises by Fréchet, Hausdorff, Menger, Moore, Sierpiński, and Alexandroff and Hopf (cf. Bibliography). Throughout the paper the symbol S will denote a basic Hausdorff space. We shall consider only spaces H, Fréchet (cf. Fréchet, (I), p. 186). Alexandroff and Hopf call such spaces T_1 spaces (loc. cit., p. 59).

DEFINITIONS. D 2.1. The symbol M_T', the *derived set of M relative to T*, means the set of all points of T which are limit points of $M\cdot T$. The symbol $\overline{M}_T$ denotes the *closure of M relative to T*; that is, $M\cdot T+(M\cdot T)_T'$.

D 2.2. If T and R are spaces (1) such that $T\supset R$ and (2) such that if M is a point set in R then $\overline{M}_R=R\cdot\overline{M}_T$, then we say that R is a *subspace* of T, that R is *embedded* in T, and that T is an *extension* of R; if, in addition, T is a Hausdorff space, it is a *Hausdorff extension* of R. If $T=\overline{R}_T$, T is an *immediate extension* of R (cf. Stone, loc. cit., p. 420).

D 2.3. If T is an immediate Hausdorff extension of S, the points of $T-S$ are called *frontier* points of S. If P is a point of T, $T\supset S$, D is an open set of S, and P does not belong to the closure in T of $S-D$, we call D an *S-neighborhood* of P. Two frontier points of S are said to *intersect* provided that each S-neighborhood of the one contains points in common with each S-neighborhood of the other.

3. **Partially ordered and quasi-partially ordered systems.** We refer the reader to a treatise by Birkhoff and an article by MacNeille for discussions, examples, and references to the literature (cf. Bibliography).

D 3.1. By a partially ordered system[6] is meant a system X in which an (*ordering*) relation "$x\leqq y$" is defined which satisfies the following conditions: (P_1) For $x\in X$, $x<x$; (P_2) if $x<y$ and $y<x$, then $x=y$; (P_3) if $x<y$ and $y<z$, then $x<z$. If the condition P_2 is omitted, then X is a *quasi-partially ordered system*. The notation "$x<y$" will be read "y includes x," "x precedes y," or "x is a part of y." If $x<y$ but not $y<x$, we say that x is a *proper* part of y.

D 3.2. If $x<y$ and $y<x$, we say that x and y are *equivalent* (*relative to the ordering* "$<$"). Note that P_2 rules out distinct equivalent elements. In this paper we shall have many examples of quasi-partially ordered systems for which P_2 does not hold.

D 3.3. An element of X is said to be *atomic*, if no element of X is a proper part of it.

D 3.4. An element of X is a *lower bound* of a subcollection of X, N, if it is a part of each element of N. The set of all lower bounds of N is called the

(6) Cf. Hausdorff, (I), p. 139 and Birkhoff, pp. 5, 7.

intersection $I(N)$ of the elements of N. A greatest element of $I(N)$ is called a *greatest lower bound* of the elements of N. The terms *upper bound* and *least upper bound* have analogous definitions. A *zero* and a *unit* of X are, respectively, a lower and an upper bound of all the elements of X.

D 3.5. If two elements of X have a lower bound in X, they are said to *intersect* (in X).

D 3.6. A subcollection of X, N, is said to be *monotonic* if for $x \in N$ and $y \in N$ either $x<y$ or $y<x$.

D 3.7. If X is partially ordered and each subcollection of X has a greatest lower bound in X, X is a *complete multiplicative system*.

D 3.8. A *lattice* is a partially ordered system X such that any two of its elements have both a greatest lower bound and a least upper bound in X; if each subset of X has such bounds, X is *complete*.

4. **The point elements.** In §1 we discussed the origin of our concept of a boundary element. This discussion applies also to a decomposition point element which satisfies (A) of D 4.1. If $M=a(E)$ is a non-isolated point of S, (4) of D 4.1 guarantees that E assist in the decomposition of M according to the theory we are to develop. For Case (A) the elements of E are subsets of $S-a(E)$; this is desirable for technical reasons in expressing relations of a point element to its *S-neighborhoods* (cf. D 4.5 and D 11.1). The fact that the topological relations of $M=a(E)$ to point sets in S depend on subsets of $S-M$ gives reasons that $S-M$ should contain S-neighborhoods of E.

D 4.1. A *boundary element of* S is a nonvacuous collection E such that (1) each of its elements is a nonvacuous, closed point set of S, (2) the product of any two elements of E contains an element of E, and (3) the product of all elements of E is vacuous. A collection of sets E is called a *point element of* S if either (A) or (B) is satisfied: (A) Either E is a boundary element of S, or there exists a non-isolated point of S, M, such that E is a boundary element of the subspace of S, $S-M$, and the following condition holds: (4) each open set in S that contains M contains an element of E[6]. (B) There exists an isolated point of S, M, which is an element of E; each element of E is a closed point set of S which contains M. If E satisfies (B) it is said to be *degenerate*.

For Case (B) E satisfies conditions (1), (2), and (4), and the product of all elements of E is M. For Case (A) the product of all elements of E is vacuous, and (4) implies (3).

D 4.2. Let $a(E)=M$ be the product of the closures in S of the elements of E. We say that $a(E)$ is *attached* to E, and conversely.

[6] The two cases under (A) may be combined as follows; this is the form of (A) we shall use in our applications: There exists an M, which is either a non-isolated point of S or is the null set, such that (1) each element of E, e, is a nonvacuous, closed point set in the space $S-M$, and its closure in S is $M+e$; (2) the product of two elements of E contains an element of E; (3) the product of all elements of E is a subset of M; (4) if M is a point of S, any open set of S which contains M contains an element of e.

D 4.3. If $a(E)$ is a point of S, E is called a *decomposition point element*. If E is a boundary element, $a(E)$ is vacuous.

D 4.4. Two point elements are said to *intersect*, provided that each element of the one has points in common with each element of the other.

D 4.5. An open set in S which contains an element of a point element E is called an *S-neighborhood* of E. A point or a frontier point of S is said to *intersect* E, provided that each S-neighborhood of the point has points in common with each S-neighborhood of E. (Cf. D 2.3.)

D 4.6. If E and F are point elements and each point element which intersects E intersects F, we say that "$E<F$"; we read this "*E is a portion of F (relative to S).*"

D 4.7. Let the aggregate of all point elements be ordered by the relation "E is a portion of F." It is easy to see that P_1 and P_3 of D 3.1 are satisfied but that P_2 is not. Thus, this system is quasi-partially ordered.

Examples. E 4.1. Let S be the Euclidean plane, M be the origin, and P_n be the point with coordinates $(1/n, 0)$. Let $e_{nk}=(P_n, P_{n+k}, P_{n+2k}, \cdots)$, and $E_{jk}=(e_{jk}, e_{j+k,k}, e_{j+2k,k}, \cdots)$. Then the E's are decomposition point elements of S and are attached to M; also, they are boundary elements of $S-M$. E_{12} and E_{22} do not intersect, but each is a portion of E_{11}. E_{33} intersects each of E_{12} and E_{22}; none of these point elements is a portion of any of the others. E_{66} is a greatest lower bound of E_{22} and E_{33}; and E_{36} is a greatest lower bound of E_{12} and E_{33}.

E 4.2. Let S be a space with infinitely many points, all of which are isolated. Then all decomposition point elements are degenerate. Any monotonic collection of point sets whose elements have a vacuous product is a boundary element.

E 4.3. Let H be a continuum in the plane whose points have coordinates (x, y) which satisfy one of the following conditions: (1) $0\leqq x\leqq 1$, $y=0$; (2) $x=1/k$ and $0<y\leqq 1/k$, where $k=1, 2, 3, \cdots$. Let F be the aggregate of all subcontinua of H which contain the origin and at least one point with a positive ordinate. Let M be the origin and E be the aggregate of all sets $f-M$, where $f\in F$. Let G be the aggregate of all sets obtained by reflecting elements of E in the X axis. Then E and G are equivalent point elements. However, *no element of either G or E is a subset of an element of the other.*

5. **The intersection of point elements; atomic elements.** Because of Theorem 5.1 the definitions which are given for the intersection of point elements in D 3.5 and D 4.4 are logically equivalent. An analogous state of affairs shows that such a consistency should not be considered obvious: Let K be a partially ordered system; define for K a new order by an application of Definitions D 3.5, D 4.6, and D 4.7; the new order need not be consistent with the basic order of K.

Theorem 5.1. *In order that two point elements E and F should intersect, it*

is necessary and sufficient that there exist a point element which is a portion of both; if this condition is satisfied and G is the aggregate of all sets which are the product of an element of E with an element of F, then (a) *G is a point element and a greatest lower bound of E and F, and* (b) $a(E)=a(F)=a(G)$.

Proof. Let H be a common portion of E and F. Since H intersects itself, it intersects each of E and F. Since E intersects H and H is a portion of F, E intersects F.

Conversely, let E and F intersect. Suppose that $a(E)$ is a point of S and $a(E)\neq a(F)$. By D 4.1, (3), there exists $f\in F$ such that f does not contain $a(E)$. Since $\bar{f}=f+a(F)$, $a(E)\in S-\bar{f}$; by D 4.1, (4), $S-\bar{f}\supset e\in E$; since $e\cdot f$ is vacuous, we are involved in a contradiction. Thus, either $a(E)$ and $a(F)$ are the same point of S, or both are the null set. If E and F are degenerate, the conclusion in (a) is obvious; cf. D 4.1, (B). Suppose, therefore, that neither is degenerate. Let $g_1=e_1\cdot f_1$ and $g_2=e_2\cdot f_2$ be elements of G, and $K=a(E)$. Then $g_1\cdot g_2=(e_1\cdot f_1)\cdot(e_2\cdot f_2)=(e_1\cdot e_2)\cdot(f_1\cdot f_2)\supset e_3\cdot f_3=g_3\in G$; thus, Condition (2) of D 4.1 is satisfied; clearly the same holds for (3) and (4). Because of (2) and (4) $\bar{g}_1\supset g_1+K$; the converse is true, since $\bar{e}_1=e_1+K$ and $\bar{f}_1=f+K$. Thus, G satisfies all the conditions of D 4.1, (A), and it is a point element. Let H be a point element which intersects G, $h\in H$, $e\in E$, and $f\in F$. Then $h\cdot(e\cdot f)$ is nonvacuous. It follows that each of $e\cdot h$ and $f\cdot h$ is nonvacuous. Thus, H intersects each of E and F, and G is a common portion of E and F.

Let X be a point element which is a common portion of E and F, and Y be a point element which intersects X. Let e, f, x, and y be elements, respectively, of E, F, X, and Y. Let Z be the aggregate of all products of an element of X by an element of Y; similarly define T in terms of E and Z. By the preceding paragraph Z is a point element and $Z<X$; since $X<E$, Z intersects E, and T is a point element. Then $T<Z<X$ and T intersects X. Since $x\cdot y\cdot e\in T$ and $X<F, f\cdot(x\cdot y\cdot e)$ is nonvacuous. Then $y\cdot(e\cdot f)$ is nonvacuous; since $e\cdot f\in G$, Y intersects G. Thus, $X<G$, and G is a greatest lower bound of E and F.

Theorem 5.2. *No decomposition point element intersects a boundary element.*

Theorem 5.3. *A degenerate point element is atomic.*

Theorem 5.4. *If M is a collection of point elements and each finite subcollection of M has a lower bound, then M has a greatest lower bound.*

This lower bound may not belong to M. The theorem has an analogy to conditions for perfect compactness; cf. Moore, (II), and Fréchet, (I), p. 231.

Proof. Let H be the aggregate of all point elements E such that E is a greatest lower bound of a finite subcollection of M; if $m\in M$, $m<m$; thus, $m\in H$ and $H\supset M$. Let K be the sum of the elements of H, and e and f be elements of K. There exist $E\in H$ and $F\in H$ such that $e\in E$ and $f\in F$. There exist two finite subcollections of M, H_E and H_F, such that E and F are greatest

lower bounds, respectively, of H_E and of H_F. By our condition $T=H_E+H_F$ has a lower bound; let β be *any* lower bound of T. Then β is a lower bound of H_E, of E, of H_F, of F, and of G, where G is a greatest lower bound of E and F (by Theorem 5.1 E and F have a greatest lower bound). Since β is *any* lower bound of T, $G<E$, and $G<F$, then G is a greatest lower bound of T; then $G\in H$ and $K\supset G$. By Theorem 5.1 $e\cdot f\in K$. Thus, Condition (2) of D 4.1 is satisfied by K. By Theorem 5.1 $a(E)=a(F)=a(G)$. It follows readily that the other conditions of D 4.1 are satisfied. Since each element of M is a subset of K, K is a portion of each of these elements. Thus, K is a lower bound of M. If the point element F is a lower bound of M, it follows from the definition of H that F is a portion of each element of H. If a point element intersects F, it intersects each element of H, and thus intersects K. Thus, $F<K$, and K is a greatest lower bound of M.

THEOREM 5.5. *If E is a point element, there exists an atomic point element which is a portion of E.*

Proof. There exists a monotonic collection of point elements, M, of which E is an element, which is not a proper subcollection of any monotonic collection of point elements; cf. Hausdorff, (I), p. 140. Let F be a greatest lower bound of M, and G be a point element such that $G<F$ (cf. Theorem 5.4). Then $G+M$ is monotonic, $G\in M$, and $F<G$; since $G<F$, by D 3.3 F is atomic.

5A. **Historical.** In §§16, 17, and 20 we discuss the space $\lambda(S)=S+M$, where S is a completely regular space, M is a collection of regular boundary points which are the atomic elements of a certain system, and $\lambda(S)$ is perfectly compact. This space has been studied by Stone, Wallman, and Čech[7]. Čech, on page 833, has a result which is equivalent to the following: In order that S be normal, it is necessary and sufficient that if F_1 and F_2 are mutually exclusive closed point sets in S, and $\beta\in\lambda(S)-S$, then β is not a limit point of both. An analogous condition on one of our ideal points β is necessary and sufficient that β be atomic[8]; this is true even if S is irregular. By Theorem 16.1 it is characteristic of a normal space that the elements of $\lambda(S)-S$ be atomic boundary points. Thus, Čech encounters the atomic points only in the case of the normal space, while we have them at our disposal for any Hausdorff space. This difference seems to be characteristic of the difference of our methods; here Čech seems to follow Tychonoff and Urysohn. Stone's methods involve an extensive use of algebra, but seem to have elements in common with those of Čech.

Wallman's methods most nearly resemble ours. He constructs the points

(7) Cf. bibliography. Stone's paper appeared in May, 1937. The author's first report before the Society was given in September, 1937; it included most of the results of the first two chapters of the present paper for the case of the normal and semi-completely normal spaces. The papers of Cartan, Wallman, and Čech appeared later.

(8) Compare the comment on Theorem 11.7 and Theorem 12.3.

of $\lambda(S)$ by means of collections of sets; for the case of the boundary points he uses what we should call atomic boundary elements. For the other type of points, those corresponding to the points of S, he uses definitions which are not suitable for our decomposition theory. Since he does not consider such a theory, his definitions are adequate for his purposes.

Cartan, on the other hand, has a basis for a decomposition theory, based on *order*, resembling that developed in our §5; however, he fails to develop the topological applications, and has nothing like the order-continuity theory of our Chapters 2 and 3. His *filter bases* and *filtres* correspond to our point elements and composition points; he proves the existence of *ultra-filtres* which correspond to our atomic points. He has the theorem: In order that two filter bases should generate the same filter, it is necessary and sufficient that each element of the one should contain an element of the other. By Example E 4.3 such a theorem is not true for our point elements. Thus, formally his definitions are not quite adequate for our treatment.

For a domain in the plane Carathéodory has a theory of chains, ends, and prime ends, which correspond to our boundary elements, boundary points, and atomic boundary points. His treatment of accessibility corresponds to our sections 6 and 10.

6. **A Theory of accessibility.** The *overlapping* of a point element and a closed set in S, which we define in D 6.1, resembles the *intersection* of two point elements (cf. D 4.4). Similarly, Theorem 6.1 is an analogue of Theorem 5.1. The discussion at the end of §5A suggests an interpretation of the results of Theorem 6.1 according to which a point element involves methods of approach or of accessibility. A boundary element may be regarded as a "way" for escaping from the basic space; the corresponding boundary point serves as a barrier for such an exit, and may be approached by such a way. In D 6.2 we formulate these ideas.

D 6.1. If E is a point element, M is a point set in S, and each element of E has points in common with M, then E and M are said to *overlap*.

D 6.2. Let E and F be point elements and M be a point set in S. If each point element which intersects E also intersects F or overlaps M, respectively, we say that *E is a way in F*, or that *E is a way in M*, respectively. Clearly, the former is equivalent to the relation "$E<F$"; the latter is analogous.

Examples. E 6.1. Let S be the plane, M be a line, Q be a point of M, and E be the aggregate of all sets $I-Q$, where I is an interval on M and Q is its midpoint. Then E is a maximal common way of M and E, and also of S and E. If R_1 and R_2 are the two rays on M from Q, then any way common to R_1 and E fails to intersect any way common to R_2 and E. We might say that there are many more ways of approach to Q in S than there are in M; and there are more in M than in R_1.

E 6.2. In the notation of Example E 4.1 the point element E_{34} is a maximal common way of the set e_{33} and the point element E_{12}.

THEOREM 6.1. *Let M be a closed point set in S and E be a point element. (1) In order that there exist a way common to M and to E, it is necessary and sufficient that M and E overlap; (2) if this condition is satisfied and $E(M)$ is the aggregate $[e\cdot M]$, where $e\in E$, then $E(M)$ is a maximal common way of E and M and $M\supset a(E(M))$.*

The content of the theorem and its proof are similar to those of Theorem 5.1. In particular, replace the "F" and the "f" of that proof by "M," and replace "G" by "$E(M)$."

THEOREM 6.2. *If F is a portion of E and G is a way common to M and F, then G is a way common to M and E.*

7. **The quasi-partial ordering of collections of point elements.** As in §4 we first define *intersection* and then define *order* in terms of intersection. The definitions of the two sections 4 and 7 are similar and the system of §4 is a subsystem of the one we consider here. In §1 we have explained how the results of §§7 and 8 afford a basis for a summation or an amalgamation theory, and how those of §4 yield a decomposition theory.

D 7.1. If a point element intersects one of a collection of point elements, the point element and the collection are said to *intersect*. Two collections of point elements are said to *intersect* if one of them intersects an element of the other.

D 7.2. Let A be the aggregate whose elements are (1) the point elements and (2) the collections of point elements. If M and N are elements of A and each element of A that intersects M also intersects N, we say that *M is a portion of N (relative to S)*, or $M<N$. Clearly, this relation is reflexive and transitive.

D 7.3. Let the relation "$M<N$" order the elements of A. Clearly D 7.3 and D 4.6 are equivalent if both M and N are point elements. Thus the ordered system of the point elements of D 4.7 is a subsystem of A.

EXAMPLE. E 7.1. Adopt the notation of E 4.1. Let $M=E_{11}$, $N=(E_{12}, E_{22})$, $K=(E_{14}, E_{24}, E_{34}, E_{44})$, and $L=(E_{14}, E_{34})$. Then M, N, and K are equivalent, and so are L and E_{12}. Also, L is a portion of M, but the converse is not true. L does not intersect E_{22}. L intersects (E_{14}, E_{24}) but neither of these is a portion of the other.

D 7.4. If M is a point element or a collection of point elements, let $\Sigma(M)$ denote the set of all point elements that are portions of M; let $\alpha(M)$ denote the set of atomic elements of $\Sigma(M)$.

THEOREM 7.1. *If M and N are collections of point elements, then M is a portion of N if and only if each point element that intersects an element of M also intersects an element of N.*

THEOREM 7.2. (1) *If N is a collection of point elements and either $M\in N$*

or $N \supset M$, *then* $M < N$; (2) *if* M *is a collection of point elements and each element of* M *is a portion of* N, *then* $M < N$; (3) $\Sigma(N) \supset N$.

Proof. Consider (2). Let $E \in M$, and F be a point element which intersects E. By Theorem 5.1 there exists a point element G which is a lower bound of E and F. Since $G < E < N$, $G < N$, by D 7.2 there exists $n \in N$ such that G and n intersect; since $G < F$, n intersects F. By Theorem 7.1 $M < N$. Clearly each element of N is a portion of N; the conclusion of (1) of the theorem follows from (2).

The following theorem shows that the ordering relation $<$ is equivalent to the aggregate-inclusion relation $\subset$, if applied to the sets $\Sigma(X)$ and $\alpha(X)$; cf. D 7.4. It brings out the importance of the atomic elements in the ordering theory.

THEOREM 7.3. *Let* M *and* N *each be point elements or collections of point elements*: (1) $M < \Sigma(M) < \alpha(M) < M$; (2) *the conditions* (a) $M < N$, (b) $\Sigma(N) \supset \Sigma(M)$, *and* (c) $\alpha(N) \supset \alpha(M)$ *are logically equivalent*; (3) *the conditions* (a) M *intersects* N, (b) *the product* $\Sigma(M) \cdot \Sigma(N)$ *is nonvacuous, and* (c) *the product* $\alpha(M) \cdot \alpha(N)$ *is nonvacuous are logically equivalent*; (4) $\Sigma(M+N) \supset \Sigma(M) + \Sigma(N)$, *and* $\alpha(M+N) = \alpha(M) + \alpha(N)$; *each of these four sets is equivalent in our ordering to* $M+N$.

Condition (3) is an analogue of Theorem 5.1. Cf., also, Theorem 7.4. Condition (3) brings out the equivalence of D 7.1 and D 3.5.

Proof. Let the point element E belong to M, and let F be a point element which intersects E. By Theorems 5.1 and 5.5 there exists an atomic point element G which is a lower bound of E and F. By Theorem 7.2 (1), $E < M$; since $G < E$, $G < M$, and $G \in \alpha(M)$. By Theorem 7.1 $M < \alpha(M)$. Since $\Sigma(M) \supset \alpha(M)$, $\alpha(M) < \Sigma(M)$; cf. Theorem 7.2 (1). By Theorem 7.2 (2), $\Sigma(M) < M$.

Let M be a portion of N and $X \in \Sigma(M)$; since $X < M < N$, $X < N$, and $X \in \Sigma(N)$; thus, $\Sigma(N) \supset \Sigma(M)$. Conversely, if $\Sigma(N) \supset \Sigma(M)$, it follows by part (1) and Theorem 7.2 (1) that $M < \Sigma(M) < \Sigma(N) < N$. Thus (2a) and (2b) are equivalent. In the same way we can show that (2a) and (2c) are equivalent.

If M and N intersect, there exist intersecting point elements E and F which belong to M and N, respectively (cf. D 7.1). By Theorems 5.1 and 5.5 there exists an atomic point element G which is a lower bound of E and F. By Theorem 7.2 (1), $E < M$ and $F < N$. It follows that $G < M$ and $G < N$, or, that $G \in \alpha(M) \cdot \alpha(N)$. Since $\Sigma(M) \cdot \Sigma(N) \supset \alpha(M) \cdot \alpha(N)$, (3a) implies each of (3b) and (3c). By D 7.1 and Theorem 5.1 each of these implies (3a).

By Theorem 7.2 (1), $\Sigma(M) \supset M$; the proof of part (4) follows with the help of this theorem and parts (1) and (2).

THEOREM 7.4. *Let* M *be a collection of point elements, and of aggregates of point elements*: (1) *The set of all the point elements which belong to elements of* M *is a least upper bound of* M; (2) *if the elements of* M *have lower a bound, the*

product of the sets $\alpha(m)$ and that of the sets $\Sigma(m)$, where m ranges over M, are greatest lower bounds of M.

Proof. Let K be the product of all the sets $\alpha(m)$ and L that of all the sets $\Sigma(m)$. By our hypothesis and Theorem 5.5 neither K nor L is vacuous. By Theorem 7.2 (2), and Theorem 7.3 (1), each of K and L is a lower bound of the elements of M. Let X be a lower bound of M, and Y be an element of M. By Theorem 7.3 $\alpha(Y) \supset \alpha(X)$; it follows that $K \supset \alpha(X)$. Since $L \supset K$, it follows from Theorems 7.2 and 7.3 that $X < \alpha(X) < K < L$. The conclusion of (2) follows. Part (1) may be proved by similar methods.

8. **The S-portions, our ideal points.** If E is a nondegenerate point element and we remove a finite number of the elements of E, the remainder is a point element which is equivalent to E. Thus, the condition P_2 for partially ordered systems does not hold for the quasi-partially ordered system of collections of point elements A, which we considered in the preceding section. In our applications to topology we find it desirable to remove the ambiguity which is involved in having equivalent, non-identical elements. We achieve this by a procedure which is essentially that of identifying equivalent elements; this yields a partially ordered system(9); its elements are our ideal points. For reasons given in §1 we call these *S-portions*.

D 8.1. Let A be the ordered system of all point elements and collections of point elements (cf. D 7.2 and D 7.3). If $E \in A$, let $P(E)$ be the collection of all elements of A, X, such that $X < E$ and $E < X$. We call $P(E)$ an *S-portion* or an *ideal point*. Clearly, $E \in P(E)$; if $X \in P(E)$, then $P(X) = P(E)$; for, if $Y \in P(X)$, $E < X < Y$, and $Y < X < E$.

D 8.2. Let P and Q be S-portions, $E \in P$, and $F \in Q$. If $E < F$ according to D 7.2, we shall say that $P < Q$, or that *P is a portion of Q*; also, we shall say that $P < F$ and $E < Q$.

D 8.3. Let $K(S)$ be the system of all S-portions; let it be ordered by the relation $P < Q$ of D 8.2. It is obvious that $K(S)$ is a partially ordered system; in particular, Condition P_2 holds.

D 8.4. If one of the elements of an S-portion is a point element, the S-portion is called a ***composition point***; if the point element is a boundary element or a decomposition point element, respectively, the point is called a ***boundary point*** or a ***decomposition point***, respectively; cf. D 4.3. The theory of composition points corresponds to that of §§4 and 5, while that of S-portions in general requires that of §7. We shall see in §14 that the composition points are included among the amalgamation points; these are a special class of the ideal points, and they are a basis for our topological applications. We shall see, also, that regular points and frontier points of S may be identified with composition points; cf. Theorem 13.7.

EXAMPLES. E 8.1. Let E be a finite collection of point elements such that

(9) Cf. Birkhoff, p. 7, Theorem 1.2.

if E_1 and E_2 are elements of E then $a(E_1)=a(E_2)$; cf. D 4.2. Then $P(E)$ is a composition point. This follows from Theorems 14.11 and 14.8; or, it may be proved directly.

E 8.2. Let P be a fixed point in the plane and R be a ray with initial point P. Let $I(n, R)$ be $I-P$, where I is the interval on R with endpoint P and length $1/n$. Let $E(R)=(I(1, R), I(2, R), \cdots)$, and let E be the set of all $E(R)$'s. Let $R_1, R_2, R_3, \cdots$ be an infinite sequence of distinct R's. Let $F_n=I(n, R_n)+I(n+1, R_{n+1})+\cdots$; let $F=(F_1, F_2, F_3, \cdots)$. Then F is a point element but does not intersect any element of E; by Theorem 8.1 $P(F)$ does not intersect $P(E)$, and *not* $P(F)<P(E)$. Suppose that H is a point element and $H<E<H$; then each $E(R_n)$ intersects H, and F intersects H; cf. D 4.4. Since $H<E$, F intersects E, and we have a contradiction. Thus, H does not exist, and $P(E)$ is *not* a composition point.

E 8.3. If E contains a boundary element and a decomposition point element, $P(E)$ is not a composition point.

D 8.5. If M is a point or a point set let $\Sigma(M)$ and $\alpha(M)$ denote, respectively, all composition points and all atomic ideal points which are portions of M (cf. D 7.4). Let E_P denote an element of the ideal point P.

D 8.6. A *degenerate* composition point is one which has degenerate elements. (Cf. D 4.1.)

THEOREM 8.1. *Let P and Q be ideal points.* (1) $\Sigma(E_P)\in P$, $\alpha(E_P)\in P$, *and* $\Sigma(E_P)\supset E_P$. (2) *The following are equivalent*: (a) $P<Q$; (b) $\Sigma(E_Q)\supset\Sigma(E_P)$; (c) $\alpha(E_Q)\supset\alpha(E_P)$. (3) *The following are equivalent*: (a) *P intersects Q*; (b) *E_P intersects E_Q*; (c) *$\alpha(E_P)\cdot\alpha(E_Q)$ is nonvacuous.* (4) *If P intersects Q, $P(\alpha(E_P)\cdot\alpha(E_Q))$ is their greatest lower bound in $K(S)$.*

(Cf. D 3.5 and the theorems of §7.)

THEOREM 8.2. *An atomic S-portion is a composition point. A composition point is atomic if and only if its elements contain only atomic point elements.*

Proof. Let $F\in E_P$ and $G\in\alpha(F)$; by Theorem 7.2 $F<E_P$. Since $G<F<E_P$, $P(G)<P(E_P)=P$. If P is atomic, $P=P(G)$, $G\in P$, and $F<E_P<G$; since G is atomic, so is F. The sufficiency may be proved by similar methods.

THEOREM 8.3. *Two atomic S-portions do not intersect.*

THEOREM 8.4. *A degenerate S-portion is atomic.*

The latter follows from Theorem 5.3. Let P and Q be atomic points which intersect in X. Then $X<P$; since P is atomic, $P=X$; similarly, $Q=X$.

9. **Intersections and orderings of collections of points.** We extend the concepts of intersection and of order to the class of all points and all collections of points. Our definitions, methods, and results are similar to those of preceding sections. We introduce the concepts of the summation and the decomposi-

tion of points. Example E 9.2 gives an indication of the complicated relations that may hold in our theory; some of the results in it may seem paradoxical.

D 9.0. The term, a *real point* is to mean a *point of S*. Hereafter, if the term *point* is used without any explicit or implicit qualifications, it may mean a real, an ideal, or a frontier point; cf. D 2.3.

D 9.1. If a real point or a frontier point intersects[10] a point element which belongs to an element of an S-portion, the point and the S-portion are said to *intersect*. Two S-portions are said to *intersect* if they have a lower bound in the system of ideal points; cf. D 3.5. If a point intersects one of a collection of points, it and the collection are said to *intersect*; similarly define the *intersection* of two collections.

D 9.2. Let B be the aggregate whose elements are the points and the collections of points. If M and N are elements of B and each ideal point that intersects M also intersects N, we say that *M is a portion of N*, or that $M<N$. Clearly, this relation gives a quasi-partial ordering for B. The formulation of our definition is similar to the condition of Theorem 7.1. By Theorem 8.1 $K(S)$ is a subsystem of B (cf. D 8.2 and D 8.3).

D 9.3. If P is a point and M is a collection of points, $P<M$, and $M<P$, we say that P is a *summation* of the elements of M; if, in addition, no two elements of M intersect, M is said to be a *decomposition* of P. $P(M)$ means an ideal point which is a summation of the elements of M.

D 9.4. If M is the set of all point elements, $P(M)$ is the *maximal* S-portion. Clearly, the term is justified, and $P(M)$ is the summation of all composition points. It follows from Theorems 8.3 and 5.5 that the set of all atomic ideal points is a decomposition of $P(M)$.

EXAMPLES. E 9.1. Let S be completely regular and R be a perfectly compact Hausdorff space which is an immediate extension of S. Then R is a decomposition of the maximal S-portion; cf. Theorem 16.1. Let M and N be the sets of all decomposition points and of all boundary points, respectively. Then $S<M$ and $M<S$; however, M and N fail to intersect. (Cf. Theorem 5.2.)

E 9.2. Let S be the plane and M be the set of all points of S. If F is a closed set, let $E(F)$ be the set of all point elements X such that each element of X is a subset of F; let $Q(F)$ be the ideal point of which $E(F)$ is an element. Let H be the aggregate of all $Q(L)$'s, where L is a line; and let K be the set of all $Q(C)$'s, where C is the circumference of a circle with a positive radius. Since no point of S is isolated, all point elements in $E(C)$ and in $E(L)$ are nondegenerate; it follows from D 4.1, Theorem 5.1 and D 7.1 that $E(C)$ and $E(L)$ do not intersect; by Theorem 8.1 $Q(C)$ and $Q(L)$ do not intersect; by D 9.1 and D 9.2 H and K do not intersect. Hence, neither $H<K$ nor $K<H$. If P is a real point common to C and L it follows from D 4.5 that P intersects elements of $E(C)$ and $E(L)$; it may be shown that P intersects $I(C)$ and $I(L)$.

---

[10] Cf. D 4.5.

Since $I(C)$ intersects L but does not intersect $I(L)$, L is not a portion of $I(L)$; similarly, C is not a portion of $I(C)$, and M is a portion of neither H nor K. By Theorem 10.3 $E(L)$ contains boundary elements, but $E(C)$ does not; it follows that $Q(C)<C$, but not $Q(L)<L$; similarly, $K<M$, but not $H<M$.

Because of the similarity of the ideas involved in §§7 and 9 we can prove theorems analogous to those of §7 for collections of points, and can use similar proofs. We give several theorems about summations and decompositions.

THEOREM 9.1. *If P is an S-portion, $\alpha(P)$ is a decomposition of P, and P is a summation of $\Sigma(P)$; $P<\alpha(P)<\Sigma(P)<P$.*

(2) *If M is a collection of ideal points, there exists one ideal point, $P(M)$, which is the summation of the elements of M; and $P(M)$ is the least upper bound of M in the system of the ideal points.*

Proof. By D 8.5 $\Sigma(P)\supset\alpha(P)$ and each element of $\Sigma(P)$ is a portion of P. By the analogue of Theorem 7.2 for the case of points, $\alpha(P)<\Sigma(P)<P$. Let X and Y be ideal points such that $Y<X$ and $Y<P$. By Theorem 8.1 $\alpha(E_P)\cdot\alpha(E_X)\supset\alpha(E_Y)$. If $F\in\alpha(E_Y)$, F is an atomic point element; by Theorems 8.1 and 8.2 $P(F)$ is atomic, $P(F)<X$, and $P(F)<P$. Hence $P(F)\in\alpha(P)$, and X intersects an element of $\alpha(P)$. By D 9.2 $P<\alpha(P)$. By Theorem 8.3 $\alpha(P)$ is a decomposition of P. This completes the proof of (1).

Consider (2). Let E be the sum of all sets $\alpha(E_m)$, where $m\in M$. Then E is a collection of point elements; let $P=P(E)$. Since $E\supset\alpha(E_m)$, by Theorem 7.2 $\alpha(E_m)<E$; by Theorem 7.3 and D 8.2 $m<P$; by the analogue of Theorem 7.2 for the case of points, $M<P$. By Theorem 8.1 if the ideal point A intersects P then E_A intersects E. By Theorems 7.3 and 7.1 E_A intersects an element of E, λ; then λ belongs to one of the sets, $\alpha(E_m)$, and E_A intersects this set; by Theorem 8.1 A intersects m; by D 9.2 $P<M$. Let the ideal point Y be an upper bound of the elements of M. Then, for $m\in M$, $\alpha(E_Y)\supset\alpha(E_m)$ (cf. Theorem 8.1). Then $\alpha(E_Y)\supset E\supset\alpha(E_m)$. By Theorem 7.3 and D 8.2 $m<P<Y$. Thus, P is a summation of the elements of M and is a least upper bound of them. Since the system of ideal points is partially ordered, P is unique.

THEOREM 9.2. *In the partially ordered system of ideal points each collection M which has a lower bound has one greatest lower bound, β_M; and $\alpha(\beta_M)$ is the product of all the sets $\alpha(m)$, where $m\in M$.*

Proof. Let Y be an ideal point which is a lower bound of the elements of M; K be the product of all the sets $\alpha(m)$; $\beta_M=P(K)$; and $m\in M$. If $\lambda\in\alpha(Y)$, then $\lambda<Y<m$, and $\lambda\in\alpha(m)$; thus $\alpha(m)\supset\alpha(Y)$, and $K\supset\alpha(Y)$. Let the ideal point X intersect Y; by Theorem 9.1 $Y<\alpha(Y)$, and by D 9.2 X intersects an element of $\alpha(Y)$; since this element belongs to K, it follows from D 9.2 that $Y<\beta_M$. Similarly, we can prove that $\beta_M<m$. By Theorem 9.1 $K=\alpha(\beta_M)$. We have established the conclusion.

10. **Accessibility and ends.** We develop a theory for the accessibility of S-portions; this involves an extension of the theory of §6, where we deal with point elements. The discussion and the examples which were given there are appropriate here. The ends may be regarded as barriers at the terminations of the corresponding ways. We borrow the term *end* from Carathéodory; in §5A we indicate similarities between his treatment and ours. Theorem 11.3 brings out a dualism of the theories of the ends and of the S-neighborhoods; both are applied in the topologization of our ideal points, and in the development of this topology. The concepts of the present section, of *overlapping* and of *being a common way*, are analogous, respectively, to those of *intersecting* and of being a *common portion*; this may be seen by comparing the results of Theorem 10.1 with those of Theorems 5.1, 7.1, 7.4, and 9.2.

D 10.1. If M is a set of real points, K is a collection of point elements, and M and some element of K overlap[11], we say that M and K *overlap*; also, $P(K)$ and M *overlap*; if $P(K)$ is a portion of the real or the frontier point Q, we say that M and Q *overlap*.

D 10.2. If P and Q are points and each S-portion that intersects P overlaps M (and intersects Q), we say that *P is an end of M (in Q)*. If P and Q are ideal points, $E \in P$, $F \in Q$, and P is an end of M in Q, we say that *E is a way to P (in F and in M)* (cf. D 6.2). Clearly, the latter implies that $P < Q$.

Examples and comment. Adopt the notation of E 4.1; the composition point $P(E_{36})$ is an end of the point set e_{33} in the point $P(E_{12})$ (cf. also E 6.2). Consider E 9.2; the point $P(F)$ may be characterized as the summation of all composition points which are ends of F, and it is an end of F; $E(F) \in P(F)$, and each point element of $E(F)$ consists of subsets of F; Theorem 10.1 shows that it is characteristic of an end of a closed set of S that it have an element E such that any point element which belongs to E consists of subsets of the closed set. Such is *not* the case for *all* the elements of an end; for, consider E 4.3; here $P(E)$ is an end of the x-axis, but no element of E is a subset of the x-axis. If, in E 9.2, we let C' and L' be the sums of all the $E(C)$'s and of all the $E(L)$'s, respectively, we have two ways in S which have no common way; the ends of these ways, $P(C')$ and $P(L')$, do not intersect.

If A is a point of a linear interval and M is a countable sequence of points on the interval which converges to A, according to our theory there exist many more methods of approach to A in the arc than there exist in the sequence. The end of the arc is to be regarded as vastly smaller than A, but is much greater than the end of the sequence; the latter is capable of further subdivision. Similar interpretations hold for the other examples.

THEOREM 10.1. *Let M be a closed point set in S and P be an S-portion.* (1) *In order that there exist an end of M in P, it is necessary and sufficient that M and P should overlap.* (2) *Let P and M overlap, $G \in P$, and $G(M)$ be the collection of all*

(11) Cf. D 6.1.

$E(M)$'s(12), *where E is a point element which belongs to G; let Q be the ideal point $P(G(M))$; then Q is the maximal end of M in P.*

Proof. Let the ideal point A be an end of M in P. Since A intersects itself, by D 10.2 A overlaps M; and $A<P$. Since A overlaps M, there exists $E_A \in A$ which overlaps M (D 10.1). Since $A<P$, by Theorem 8.1 $\Sigma(E_P) \supset E_A$; by D 10.1 M overlaps each of $\Sigma(E_P)$ and P. The necessity of the condition in (1) follows.

Conversely, suppose that P and M overlap; by D 10.1 there exist $G \in P$ and a point element X, which belongs to G, such that X overlaps M; by Theorem 6.1 $X(M)$ is a point element; thus, $G(M)$ and $Q=P(G(M))$ exist. Let Z be an ideal point which intersects Q. By Theorem 8.1 E_Z intersects $G(M)$; by D 7.1 there exist point elements E_1 and $E(M)$ which belong to E_Z and to $G(M)$, respectively, and which intersect. Since each element of $E(M)$ is a subset of M, it follows by D 4.4 that each element of E_1 has points in common with M. By D 6.1 and D 10.1 M overlaps each of E_1, E_Z, and Z. There exists an element of G, E, such that each element of $E(M)$ is a subset of an element of E; then E_1 intersects E; by D 7.1 E_Z intersects G; by Theorem 8.1 Z intersects P. Since, also, Z overlaps M, Q is an end of M in P (cf. D 10.2).

Let G_1 be any element of P, whatever. Since each element of $E(M)$ is a subset of an element of E, it follows from D 4.4 and D 4.6 that $E(M)<E$; by Theorem 7.2 $G(M)<G$, and thus $G(M)<G_1$; by Theorem 7.1 $E(M)$ intersects an element, K, of G_1; since each element of $E(M)$ is a subset of M, it follows that K overlaps M. By the results of the preceding paragraph it follows that $G_1(M)$ is nonvacuous; and, if $Q_1=P(G_1(M))$, then Q_1 is an end of M in P.

Let β be an end of M in P, and Y be an ideal point which intersects β. Then $\beta<P$; by Theorem 8.1 $\alpha(G)\supset\alpha(E_\beta)$, and $K=\alpha(E_\beta)\cdot\alpha(E_Y)$ is nonvacuous. Let $F \in K$; by Theorems 7.3 and 7.2 $F<E_\beta$; by D 8.2 $P(F)<\beta$; by D 10.2 $P(F)$ overlaps M; by the result of the preceding paragraph, F overlaps M; by Theorem 6.1 $F(M)$ exists; by arguments we have used before, $F(M)<F<E_\beta<G$. By D 7.2 $F(M)$ intersects an element E of G. Since each element of $F(M)$ is a subset of M, it follows by an argument we used in the next-to-the-last paragraph that $E(M)$ exists and $F(M)$ and $E(M)$ intersect. Since $\alpha(E_Y)\supset K$, $F\in K$, and $F(M)<F$, it follows by Theorems 7.2 and 7.3 that $F(M)<F<K<E_Y$, and that $F(M)\in\Sigma(E_Y)$. Thus, $\Sigma(E_Y)$ and $G(M)$ intersect, since their respective elements $F(M)$ and $E(M)$ intersect. By Theorem 8.1 Y and Q intersect. Thus, by D 9.2 $\beta<Q$. Since β is an end of M in P, Q is the maximal end of M in P. Since $Q_1<Q$, and $Q<Q_1$, $Q=Q_1$. Thus, Q is independent of the particular element G of P. We have established the truth of (2) and the sufficiency of the condition in (1).

THEOREM 10.2. *Let M be a closed point set in S, P be an end of M, and Q*

(12) Cf. Theorem 6.1. For example, $E(M)$ is the set of all products $e\cdot M$, where $e \in E$.

be an ideal point which intersects P; there exists an end of M which is common to P and Q. If $Q<P$, Q is an end of M.

THEOREM 10.3. *In order that the closed point set in S, M, be perfectly compact in itself, it is necessary and sufficient that no boundary point be an end of M.*

Proof. Suppose that M is not perfectly compact in itself; there exists in M a monotonic collection of closed point sets, K, which have no point in common (cf. Moore, (II)). By D 4.1, K is a boundary element. By Theorem 10.1, $P(K)$ is an end of M.

Conversely, let M be perfectly compact in itself, and E be a boundary element. Let K be the aggregate of all products $e\cdot M$, where $e\in E$. If E overlapped M, no element of K would be vacuous; then the product of the elements of K would be nonvacuous (cf. Fréchet, (I), p. 231). Then the product of the elements of E would be nonvacuous; by D 4.1 this is impossible. Since E does not overlap M, there exists no end of M in $P(E)$ (cf. Theorem 10.1).

THEOREM 10.4. *If the point set M of S is perfectly compact in itself and P is a point and a limit point of M, there exists a composition point which is an end of M in P.*

Proof. If R is a neighborhood of P relative to M, let $e(R)=\overline{R}_M-P$. Let E be the aggregate of all the $e(R)$'s. Since M is a closed point set in S, $\overline{R}_M$ is closed in S, and $e(R)$ is closed in $S-P$. Let U be an open set in S which contains P; then $U\cdot M$ is a neighborhood of P relative to M. M may be regarded as a regular subspace of S (cf. Alexandroff and Hopf, p. 89, Theorem IX). It follows that $U\cdot M$ and U contain an element of E. If we let $P=a(E)$, E satisfies condition (4) of D 4.1. It may readily be verified that E satisfies the remaining conditions. Thus, E is a point element, and $Q=P(E)$ is a composition point. By Theorem 10.1 Q is an end of M.

Let the ideal point X intersect Q; by D 10.2 X overlaps M. By Theorem 8.1 E_X intersects E; by D 7.1 there exists a point element, F, which belongs to E_X and intersects E. Since each S-neighborhood of P contains an element of E, and each element of F has points in common with each element of E, it follows that F and P intersect; cf. D 4.5. By D 9.1 X intersects P; by D 10.2 Q is an end of M in P.

11. **The S-neighborhoods.** These are applied in the topologization of our ideal points. The condition of "being separated by S-neighborhoods" of D 11.2 is analogous to the separations which are involved in Hausdorff's Axiom D and in the definition of normality; it finds analogous applications; cf. Theorem 11.5. The end theory of the preceding section is closely related to the S-neighborhood theory; duality of the two is brought out in Theorem 11.3. An important characterization of the atomic points is given in Theorem 11.7.

D 11.1. An open set in S, D, is called an *S-neighborhood of M* if one of the following conditions holds: (1) M is a point or a subset of D; (2) M is a point

element and D contains an element of M; (3) M is an ideal point, a collection of points, or an aggregate of point elements, and D is an S-neighborhood of each element of M. Thus, to see whether D is an S-neighborhood of the S-portion P, first apply (2) for each point element which belongs to some element of P, then apply (3) for each element of P; then apply (3) for P. (Cf. D 2.3.)

D 11.2. If U and V are mutually exclusive open sets in S which are S-neighborhoods of M and of N, respectively, we say that they *separate* M *and* N (*relative to* S).

Theorem 11.1. *If K is an S-portion or a collection of point elements, and D is an open set in S, then D is an S-neighborhood of K if and only if K and $S-D$ do not overlap.*

Proof. Let D be an S-neighborhood of the S-portion P, $E\in P$, and $\lambda\in E$. By D 11.1 D is an S-neighborhood of λ and of E; and there exists $\beta\in\lambda$ such that $D\supset\beta$. Then $\beta\cdot(S-D)$ is vacuous; by D 6.1, Theorem 10.1, and D 10.1, neither λ, E, nor P overlaps M. The converse may be established by reversing this argument.

Theorem 11.2. *If E is an element of the ideal point P, then E and P have the same S-neighborhoods.*

Proof. Let D be an open set in S. By Theorem 10.1 if some element of P overlaps $S-D$, so does every other element. The conclusion follows from Theorem 11.1 and D 11.1.

Theorem 11.3. *Let D be an open set in S and P be an ideal point.* (1) *In order that D be an S-neighborhood of P it is necessary and sufficient that $S-D$ have no end in P*; (2) *in order that P be an end of $S-D$ it is necessary and sufficient that D fail to be an S-neighborhood of any portion of P.*

Proof. If D is an S-neighborhood of P, by Theorem 11.1 P does not overlap $S-D$. By Theorem 10.1, $S-D$ has no end in P. A similar argument proves the converse.

Let P be an end of $S-D$ and X be a portion of P. Since X intersects P, by D 10.2 X overlaps $S-D$; by Theorem 11.1 D is not an S-neighborhood of X. Conversely, suppose that D is not an S-neighborhood of any portion of P. Let Z be an ideal point which intersects P. By Theorem 8.1 $P(\alpha(E_P)\cdot\alpha(E_Z))$ is a lower bound of P and Z. By our hypothesis and Theorem 11.1 $S-D$ overlaps $P(\alpha(E_P)\cdot\alpha(E_Z))$; by Theorem 10.1 and D 10.1 $S-D$ overlaps $\alpha(E_P)\cdot\alpha(E_Z)$, $\alpha(E_Z)$, and Z. By D 10.2 P is an end of $S-D$.

Theorem 11.4. *Let D be an S-neighborhood of the ideal point X, and P be a non-isolated point of D; then $D-P$ is an S-neighborhood of X.*

Proof. If E is a point element which belongs to an element of X, by D 11.1 there exists $e_1\in E$ such that $D\supset e_1$. If $a(E)=P$, it follows from the fact that P

is non-isolated that $S-P\supset e_1$ (cf. D 4.1 and D 4.2). If $a(E)\neq P$ there exist elements of E, e_2 and e, such that $S-P\supset e_2$ and $e_1\cdot e_2\supset e$ (cf. D 4.1). Then $D-P\supset e$. By D 11.1 $D-P$ is an S-neighborhood of E and of X.

THEOREM 11.5. *In order that a real or a frontier point should fail to intersect a composition point, it is necessary and sufficient that the two points be separated by a pair of their respective S-neighborhoods.*

Proof. Let the real point P be separated from the composition point Q by U and V. Then by D 11.1 V is an S-neighborhood of any point element which belongs to an element of Q; by D 4.5 and D 9.1 P intersects neither the point element nor Q. Conversely, suppose that P and Q do not intersect, and let the point element E be an element of Q. By D 9.1 P and E do not intersect; by D 4.5 P and E can be separated by a pair of their respective S-neighborhoods, U and V; by Theorem 11.2 V is an S-neighborhood of the point Q.

THEOREM 11.6. *If each of M and N is an ideal point or a collection of ideal points and $M<N$, then each S-neighborhood of N is an S-neighborhood of M.*

Proof. Suppose that D is an S-neighborhood of N but not of M; there exists $P\in M$ such that D is not an S-neighborhood of P. By Theorem 11.3 $S-D$ has an end in P, say Q; then $Q<P$. Since Q intersects P and $M<N$, by D 9.2 Q intersects an element Z of N; let β be a lower bound of Q and Z. If $\lambda\in\beta$, $\Sigma(E_Z)\supset\lambda$ and $\Sigma(E_Z)\in Z$ (cf. Theorem 8.1). Since $Z\in N$, D is an S-neighborhood of Z, of $\Sigma(E_Z)$ and of λ (cf. D 11.1). Since λ is an arbitrary element of β, D is an S-neighborhood of β; since $\beta<Q$, by Theorem 11.3, (2), we are involved in a contradiction. Thus, D does not exist.

THEOREM 11.7. *An S-portion P is atomic if and only if it satisfies the following condition: If the sum of a finite number of open sets in S is an S-neighborhood of P, at least one of them is an S-neighborhood of P.*

If P is nondegenerate, the following is an equivalent condition: *If M and N are closed point sets in S and have at most a finite number of common points, then P is not an end of both M and N.*

Proof. Let D_1 and D_2 be open sets whose sum D is an S-neighborhood of P. Let $S-D_1=M_1$ and $S-D_2=M_2$. Suppose that neither D_1 nor D_2 is an S-neighborhood of P; by Theorem 11.3 M_1 and M_2 have ends in P. If P is atomic, these ends are P. Let E be a point element such that $E\in P$. There exists $e\in E$ such that $D\supset e$. By Theorems 6.1 and 10.1 $E(M_1)$ and $E(M_2)$ are point elements and are elements of P. Then $e\cdot M_1\in E(M_1)$ and $e\cdot M_2\in E(M_2)$. Since any two elements of P intersect, by D 4.4 $(e\cdot M_1)\cdot(e\cdot M_2)$ is nonvacuous. Since $D\supset e$ and $M_1\cdot M_2=S-D$, we have a contradiction. Thus, the condition is necessary.

Suppose that P is not atomic and D is an S-neighborhood of P. By Theo-

rems 8.1, 8.2, and 8.3 there exist two atomic point elements, E and F, which belong to $\alpha(E_P)$ but do not intersect. By D 4.4 and D 11.1 there exist $e\in E$ and $f\in F$ such that $D\supset e+f$ and $e\cdot f$ is vacuous. Let $D_1=D-\bar{e}$ and $D_2=D-\bar{f}$. Since D_1 does not contain e, it is not an S-neighborhood of E or of P; similarly, D_2 is not an S-neighborhood of P. By D 4.1 and D 4.2 if $a(E)\neq a(F)$, then $\bar{e}\cdot\bar{f}$ is vacuous and $D_1+D_2=D$. By D 4.1 if $a(E)=a(F)$, then $e+a(F)=\bar{e}$, $f+a(F)=\bar{f}$, and $\bar{e}\cdot\bar{f}=a(F)$; since $e\cdot f$ is vacuous, $a(F)$ is non-isolated. Then $D_1+D_2=D-a(F)$, and by Theorem 11.4 this sum is an S-neighborhood of P. Thus, the condition is sufficient.

Chapter II. Topological properties of the ideal points

In the first chapter we considered intersection and order; our treatment now requires the introduction of relations which involve continuity. In §12 we introduce a topology with the help of the S-neighborhoods. As a suggestion of problems we shall consider and of methods we shall use, we refer the reader to Alexandroff and Hopf, loc. cit., pp. 95–98; note particularly Theorem VIII. In this theorem they consider a perfectly compact Hausdorff space X, a decomposition $\sum A$ of X into mutually exclusive closed point sets, and a mapping $y=\alpha(x)$ of X into a space Y such that $\alpha(x_1)=\alpha(x_2)$ if and only if x_1 and x_2 are elements of the same element of the decomposition. They develop equivalent conditions, which include: (1) the transformation is continuous and Y is a perfectly compact Hausdorff space; (2) the transformation is continuous, and Y is a Hausdorff space; (3) the decomposition is upper semi-continuous.

In our Theorems 15.3, 15.4 and 15.5 we obtain extensive generalizations of the results of these authors. These results involve properties which restrict our attention to a special class of the S-portions, the amalgamation points. In §14 we develop a geometric theory of these points in which the amalgamation points have properties similar to those of the perfectly compact point sets in Alexandroff and Hopf's treatment. In Theorem 15.5 this analogy is strengthened by a kind of representation theory([18]) in which the amalgamation points are identified with certain perfectly compact collections of amalgamation points, to which Alexandroff and Hopf's results may be applied. Corresponding to the set-theoretic treatment of upper semi-continuity of these authors our treatment requires the introduction of a geometric upper semi-continuity theory; cf. D 15.1. The latter part of the paper is devoted to the application of the basic theorems of §15.

In §13 we give characterizations of various types of regular and normal spaces in terms of the regular composition points. In §§16 and 17 we find that these characterizations prove useful in applying the results of §15. Such applications are particularly simple and extensive for the case of the semi-completely normal spaces, which we introduce in §13. For them the development

([18]) Cf. Birkhoff, p. 76. Our representations involve not merely order but also continuity.

of Chapter II has a completeness that is attained for more general spaces only in Chapter III.

12. **The topology of our points.** We first give the following definition.

D 12.1. If P is a point and M is a point set, and each S-neighborhood of P is an S-neighborhood of infinitely many elements of M, then P is a *limit point* of M. Similarly, if P and M are a point element and a collection of point elements, respectively, P is called a *limit element* of M.

THEOREM 12.1. *The space $T(M)$, which is obtained by topologizing M by* D 12.1 *is a space H Fréchet. Let U be an open set in S and let the symbol $U(M)$ denote any nonvacuous subset of M consisting of all elements of M, with the possible exception of a finite number, of which U is an S-neighborhood. The collection of all such $U(M)$'s may be taken as a system of neighborhoods for defining limit points in $T(M)$.*

If P is the point and N is a subset of M, clearly P is a limit of N according to D 12.1 if and only if each $U(M)$ that contains P contains points of N distinct from P. By Fréchet, (I), pp. 185–186, $T(M)$ is a space H.

THEOREM 12.2. *If the composition point P neither intersects nor is a limit point of the set M of composition points, there exists an S-neighborhood of P which is not an S-neighborhood of any element of M.*

Proof. P has an S-neighborhood D, which is an S-neighborhood of at most a finite number of elements of M, $X_1, X_2, \cdots, X_m$. Let E and F_i be point elements which are elements of P and of X_i, respectively. There exist $e_i \in E$ and $f_i \in F_i$ such that $e_i \cdot f_i$ is vacuous. By D 4.1 there exists $e \in E$ such that $e_1 \cdot e_2 \cdot e_3 \cdots e_m \supset e$. Let $U = (S - \bar{f}_1) \cdot (S - \bar{f}_2) \cdots (S - \bar{f}_m)$. Then $U \supset e$ and U is an S-neighborhood of E and of P; but U is not an S-neighborhood of F_i or of X_i. Then $D \cdot U$ satisfies the conclusion.

THEOREM 12.3. *In order that a composition point be a limit point of a closed set of S, it is necessary and sufficient that the point be nondegenerate and contain an end of the set.*

Proof. Let M be the set, P be a point which satisfies the condition, and E be a point element which is an element of P. By Theorem 10.1 P and E overlap M; by D 6.1 each element of E contains points of M. By the hypothesis and D 8.6 E is nondegenerate; by D 4.1 no point of M is common to all elements of E; thus, each of these elements contains infinitely many points of M; by D 11.1 each S-neighborhood of P contains an element of E. Thus, P is a limit point of M.

If there exists no end of M in P, by Theorem 11.3 $S-M$ is an S-neighborhood of P. Then P is not a limit point of M. If P is degenerate by D 8.6 and D 4.1 $a(E)$ is an isolated point of S; then $a(E)$ is an S-neighborhood of E and of P, and P is not a limit point of M.

D 12.2. If M is a collection of point elements, let $a(M)$ be the collection of all real points which are attached to elements of M (cf. D 4.2). If M is a real point or a frontier point, let $a(M)$ be M or the null set, respectively. If E is a set of points, or is an ideal point, let $a(M)$ be the sum of all sets $a(m)$, where $m \in M$. The concept $a(M)$ is applied in the definition of regularity (cf. D 13.1). In Example E 9.2, $a(Q(F))=F$. The method of the definition involves an induction similar to that of D 11.1.

THEOREM 12.4. *Let P and Q be ideal points, $P<Q$, $E \in P$, and D be an S-neighborhood of P. Then $a(Q) \supset a(P)$, $a(P)=a(E)$, and $\overline{D}_S \supset a(P)$.*

Proof. By D 8.2 if $F \in Q$, $E<F$. If $X \in a(E)$, there exists $\lambda \in E$ such that $a(\lambda)=X$. By Theorem 7.1 there exists $\delta \in F$ such that λ intersects δ; by Theorem 5.1 $a(\lambda)=a(\delta)$. Thus, $a(F) \supset a(E)$. Since $a(Q)=\sum a(F)$, for $F \in Q$, and $a(P)=\sum a(E)$, for $E \in P$, it follows that $a(Q) \supset a(P)$. If $P=Q$, then $E<F$, and $F<E$; then $a(E) \supset a(F)$, and $a(F) \supset a(E)$; thus, $a(E)=a(F)$; since F is any element of $Q=P$, $a(E)=a(P)$. By D 11.1 D is an S-neighborhood of $\lambda \in E \in P$. If $X \in S-\overline{D}_S$, by D 4.1 and D 12.2 X is not an element of $a(\lambda)$, of $a(E)$, or of $a(P)$.

13. **Regular points and normal spaces.** Since we deal extensively with perfectly compact Hausdorff spaces of our ideal points, we find need of the concept "the regularity relative to S of an ideal point." In terms of this concept we characterize various types of regular and of normal spaces (Theorems 13.1 to 13.3). These characterizations are of interest, since the regularity of a Hausdorff space does not imply its normality. These characterizations yield methods for applications of the decomposition and mapping theory of §15; the most extensive results are given for the semi-completely normal spaces, which we introduce in this section. Normality proves to be a weaker condition than semi-complete normality; and the latter is weaker than complete normality. In Theorem 13.7 we show that regular real or frontier points may be replaced by regular composition points in questions which involve either continuity or order. This procedure gives a basis for applying our theory to questions concerning the embedding of S in other spaces, and to those involving mappings of S. In general our methods do not require the regularity or the normality of the basic space; they are applicable locally, so long as they deal with regular ideal points. For instance, we are in a position to study an infinitesimal regular portion of an irregular point. For the plane all composition points are regular.

D 13.1. A point P is *regular* (*relative to* S) provided that if D is an S-neighborhood of P there exists an S-neighborhood of P, R, such that $D+a(P) \supset \overline{R}_S$ (cf. D 12.2 and Alexandroff and Hopf, p. 68).

D 13.2. S is said to be *semi-completely normal* provided that if P is a point of S, then the subspace of S, $S-P$, is normal.

Clearly each completely normal space has this property; the plane is an

example. In Theorem 13.4 we show that each semi-completely normal space is normal; an example by Tietze shows that the converse is not true (loc. cit., pp. 304–306). An example by F. B. Jones shows that *if the hypothesis of the continuum is valid*, then there exists a semi-completely normal space which is not completely normal. He has constructed a space which is normal (loc. cit., pp. 673–675). On page 674 he has a point set N, which is uncountable and contains no limit point of itself. By his Theorem 4 and the *continuum hypothesis* S is not completely normal. It is obvious from his construction of S that if P is a point of S then $S-P$ is normal.

THEOREM 13.1. (1) *In order that a point of S be regular, it is necessary and sufficient that it intersect no boundary point.* (2) *An irregular point of S intersects an irregular boundary point.*

Thus, a regular space is characterized by the property that it be a Hausdorff space and that none of its points intersect a boundary point.

Proof. Let P be irregular, and U be an open set containing P such that if R is an S-neighborhood of P, then $\bar{R}\cdot(S-U)$ is nonvacuous. If $M=\bar{U}-U$ were perfectly compact, it would be possible to separate P and M by a pair of open sets, U_1 and U_2(14); then $U\cdot U_1$ is an open set which contains P and is a subset of U; the closure of this set is a subset of U, since it contains no point of U_2 and $U_2\supset M$. Thus we are involved in a contradiction; it follows that there exists a collection of open sets, G, which covers M such that no finite subcollection of G covers M (cf. Kuratowski and Sierpiński, loc. cit.). For $X\in M$ let $U(X)$ and $W(X)$ be a definite pair of mutually exclusive open sets which contain X and P, respectively, such that $U(X)$ is a subset of some element of G. Clearly, no finite collection of the $U(X)$'s covers M. Let D denote the sum of a finite collection of the $U(X)$'s, $F=M-M\cdot D$, and E be the set of all F's. It follows that E is a boundary element; let it be an element of the boundary point β. Suppose that β is regular; by Theorem 11.4 $S-P$ is an S-neighborhood of β; since $a(\beta)$ is vacuous, by D 13.1 there exists an S-neighborhood of β, V, such that $S-P\supset\bar{V}$; then $S-\bar{V}$ is an open set containing P. Since V contains an element of E, $M-M\cdot V$ can be covered by a finite collection of the $U(X)$'s, say $U(X_1)$, $U(X_2)$, $\cdots$, $U(X_n)$. The sum of V and this finite collection is an open set which contains M. Let $\delta=U\cdot(S-\bar{V})\cdot W(X_1)\cdots W(X_n)$. Then δ is an open set which contains P, $\delta\cdot M$ is vacuous, and $U\supset\bar{\delta}$. Thus, the supposition that β is regular involves a contradiction. Suppose that P and β did not intersect; then by Theorem 11.5 β and P can be separated by a pair of their respective S-neighborhoods, V and W_1; by Theorem 11.2 and D 11.1 V contains an element of E; since $\bar{V}$ does not contain P we arrive at a contradiction as we did above. Thus, we have established the truth of (2) and the sufficiency of the condition in (1).

(14) Follow the proof of Theorem IX, p. 89, Alexandroff and Hopf.

Let P be regular, E be a boundary element, and $\beta = P(E)$. By D 4.1 there exists $e \in E$ such that $P \in S - e$. Since e is closed in S, there exists an open set U, containing P such that $S - e \supset \overline{U}$. By D 11.1 and Theorem 11.2 $S - \overline{U}$ is an S-neighborhood of E and of β. By Theorem 11.5 P and β do not intersect. Thus, the condition is necessary.

THEOREM 13.2. *A Hausdorff space is normal if and only if each of its boundary points is regular.*

Proof. Let S be normal, E be a boundary element, and D be an S-neighborhood of the boundary point $P(E)$. By D 11.1 there exists $e \in E$ such that $D \supset e$. Since $S - D$ and e are closed and mutually exclusive, they can be separated by a pair of mutually exclusive open sets U and V. Since $V \supset e$, by Theorem 11.2 V is an S-neighborhood of $P(E)$; also $D \supset \overline{V}$.

Conversely, let the condition hold in S and M and N be mutually exclusive closed sets. By Theorem 13.1 each point of S is regular. For $P \in M$ let $D(P)$ be an open set which contains P such that $N \cdot \overline{D(P)}$ is vacuous. Let D denote any open set which is the sum of a finite number of the $D(P)$'s; and let E be the aggregate of all the sets $M - M \cdot D$. If one of the D's contains M, the complement of its closure contains N; if not, E is a boundary element. Since $S - N \supset M$, $S - N$ is an S-neighborhood of $P(E)$; by our condition there exists an S-neighborhood of $P(E)$, U, such that $S - N \supset \overline{U}$; then U contains an element of E, and there exists a D such that $U \supset M - M \cdot D$. Then $D + U \supset M$, and $N \cdot (\overline{D} + \overline{U})$ is vacuous. Thus, the complement of $\overline{D} + \overline{U}$ is an open set which contains N, and the condition is sufficient.

THEOREM 13.3. *A Hausdorff space is semi-completely normal if and only if each of its composition points is regular.*

Proof. Let S be semi-completely normal, β be a composition point, and E be a point element which is an element of β. If β is degenerate, then $a(\beta)$ is an isolated point of S and is an S-neighborhood of β; then β is regular (cf. D 8.6 and D 4.1). If $a(\beta)$ is a non-isolated point of S, then by D 13.2 the subspace $S - a(\beta)$ of S is normal; then β is a boundary point of this subspace; by Theorem 13.2 β is a regular boundary point of this subspace. It follows with the help of Theorem 11.4 that β is a regular decomposition point of S. If β is a boundary point of S let E be a point element, $E \in \beta$; let P be a point of S; and let F be the collection of all elements of E which do not contain P; it follows that F is a boundary element of S and of $S - P$; also, in S, $F < E$, $E < F$, and $F \in \beta$; by the argument we used above it follows that F is an element of a regular boundary point of $S - P$; and, hence, that β is regular relative to S.

Conversely, let all composition points of S be regular and let P be a point of S. By Theorem 13.2 S is normal and $S - P$ is regular. Methods analogous to those used for Theorem 13.2 show that $S - P$ is normal; by D 13.2 S is semi-completely normal.

THEOREM 13.4. *A semi-completely normal space is normal.*

By Theorem 13.3 all the boundary points are regular; the conclusion follows from Theorem 13.2.

THEOREM 13.5. *If F is a finite point set in a semi-completely normal space S, then $S-F$ is semi-completely normal.*

This may be proved with the help of Theorem 13.3 and the methods used in establishing that theorem.

THEOREM 13.6. *Let a regular point P be added to S so that the space $S+P$ is a Hausdorff space. If S is normal, so is $S+P$; in order that $S+P$ be semi-completely normal, it is necessary and sufficient that S have the same property.*

Thus a semi-completely normal space is characterized by the following property: If we either add or remove a finite number of regular points, or do both, and the result is a Hausdorff space, it is normal.

Proof. Let Q be a boundary point of the space $S+P$. Since P is a regular point, by Theorem 13.1 Q and P do not intersect relative to $S+P$; by Theorem 11.5 there exist in $S+P$ mutually exclusive open sets U and V which are $(S+P)$-neighborhoods of Q and P, respectively. Let E be a point element of $S+P$ such that no element of E contains P and that $E \in Q$; then E is a point element of S. Since S is normal and U is also an S-neighborhood of E, it follows from Theorem 13.2 that there exists an S-neighborhood of E, W, such that $U \supset \overline{W}$. Since the closures of W in S and in $S+P$ are identical, it follows that Q is a regular boundary point of $S+P$. By Theorem 13.2 $S+P$ is normal.

THEOREM 13.7. *Let P be a real, a frontier, or a composition point which is regular; if $a(P)$ is a non-isolated point of S, let $F(P)$ be the aggregate of all sets, $\overline{D}_S - a(P)$, where D is an S-neighborhood of P; otherwise, let $F(P)$ be the set of all the $\overline{D}_S$'s. Let $Q(P)$ be the composition point of which $F(P)$ is an element. Then* (1) $a(P)=a(Q(P))$, $P<Q(P)$, *and* $Q(P)<P$; (2) *if P is a limit point of a point set, then so is $Q(P)$; and conversely*; (3) *each S-neighborhood of P is an S-neighborhood of $Q(P)$; if D is an S-neighborhood of $Q(P)$, either D or $D+a(P)$ is an S-neighborhood of P*; (4) *$Q(P)$ is regular.*

Thus, $Q(P)$ is equivalent to P both from the point of view of order and of continuity. The topological applications of our theory require such a twofold equivalence; we discuss this in the introductions to Chapter II and to §§14 and 15.

Proof. Let P be a composition point and E be a point element such that $E \in P$. Since each element of $F(P)$ contains an element of E, $E<F(P)$ (cf. D 4.6). Let H be a point element which does not intersect E; there exist $e \in E$ and $h \in H$ such that $e \cdot h$ is vacuous. Then $S-\bar{h} \supset e$; by Theorem 11.2 $S-\bar{h}$

is an S-neighborhood of P. By D 13.1 there exists an S-neighborhood of P, U, such that $a(P)+(S-\bar{h})\supset\overline{U}_S$. Then $\overline{U}_S-a(P)$ or $\overline{U}_S$ is an element of $F(P)$, and contains no point of h. By D 4.4 $F(P)$ and H do not intersect. (It may readily be verified by D 4.1 that $F(P)$ is a point element.) It follows that $F(P)<E$ (cf. D 4.6). Since also $E<F(P)$, it follows by D 8.2 that $P<Q(P)$ and $Q(P)<P$; thus, $P=Q(P)$.

Next, let P be a real point or a frontier point. Clearly, $F(P)$ satisfies the conditions of D 4.1 and is a point element. Let X be a composition point which does not intersect P. By Theorem 11.5 P and X can be separated by a pair of their respective S-neighborhoods, U and V. There exists an S-neighborhood of P, W, such that $U\supset\overline{W}_S$. Then U contains an element of $F(P)$. By D 11.1 V contains elements of each point element which belongs to X; by D 4.4, D 7.1, and Theorem 8.1 $Q(P)$ and X do not intersect. By D 9.2 $Q(P)<P$. Conversely, let X be an ideal point which does not intersect $Q(P)$; by Theorem 8.1, if G is a point element belonging to X, G does not intersect $F(P)$. By D 4.4 there exist $f\in F(P)$ and $g\in G$ such that $f\cdot g$ is vacuous. There exists an S-neighborhood of P, D, such that either $f=\overline{D}_S$ or $f=\overline{D}_S-P$. By D 4.1 $S-\overline{D}_S$ contains g and is an S-neighborhood of G. By D 4.5 and D 9.1 P intersects neither G nor X. By D 9.2 $P<Q(P)$.

Since P is regular, each of its S-neighborhoods contains an element of $F(P)$ and is an S-neighborhood of $Q(P)$. If D is an S-neighborhood of $Q(P)$, $D\supset f\in F(P)$ (D 11.1 and Theorem 11.2). Then $f+a(P)=\overline{R}_S$, where R is an S-neighborhood of P and of $Q(P)$. Then $D+a(Q(P))\supset\overline{R}_S\supset R$, and $Q(P)$ is regular; since $a(Q(P))=a(P)$, either $D+a(P)$ or D is an open set containing R, and is an S-neighborhood of P.

If P is a non-isolated point of S, the conclusion of (2) follows from (3), (4), D 12.1, and Theorem 11.4. For the other cases, P and $Q(P)$ have the same S-neighborhoods, and (2) follows.

14. **The amalgamation points.** Suppose that the regular space T is an immediate extension of S. By Theorem 13.7 we can replace each point of S by an equivalent decomposition point, and each point of $T-S$ by a boundary point, and thus obtain a space T_1 of composition points which is topologically equivalent to T. We now introduce the amalgamation points in order to make more general applications, such as those we discuss in the introductions of Chapter II and of §13. These applications require that the amalgamation points include the composition points and that they have special properties which do not hold for all ideal points. These properties are similar to those of closed point sets in perfectly compact Hausdorff spaces. Thus, in Theorems 14.3, 14.4, and 14.5 we establish for amalgamation points a property analogous to the Borel-Lebesgue covering property; this analogy is emphasized in Theorem 15.3. Theorem 14.11 corresponds to the theorem that a nonvacuous product of perfectly compact sets is perfectly compact. In Theorems 14.6 and 14.7 we establish a condition for the separation of non-intersecting amalgama-

tion points; this suggests the condition for the separation of closed sets, which characterizes the normality of a Hausdorff space.

In Theorems 14.12 and 14.13 we add to the characterizations of the normal and the semi-completely normal spaces. For such spaces we obtain an extensive knowledge concerning the atomic elements of the developments of §§15 and 16. For the Hausdorff spaces in general such a completeness is attained in Chapter III only after the development of a theory of multiplicative systems and lattices of regular amalgamation points. Theorems 14.10 and 14.11 afford a basic technique for this lattice theory. The examples at the end of this section give indications of the extent to which our results may find applications in irregular spaces.

D 14.1. An S-portion, P which satisfies the following condition is called an *amalgamation point*: If $E \in P$ and X is a point element which is a limit element[18] of E, then X and E intersect. In the first chapter we defined *order* in terms of *intersection*. Since both order and continuity relations are used in our topological applications of the amalgamation points, it may be noted that the definition involves both order and continuity.

D 14.2. If M is a collection of points, P is an amalgamation point, $P<M$, and $M<P$, we say that P is the *amalgamation* of the elements of M. That is, the *summation* of D 9.3 is called an *amalgamation* if the result of the summation process is an amalgamation point.

Examples. E 14.1. Let S be the plane, E be the set of all decomposition point elements, and F be the set of all boundary elements. By Theorems 13.7, 13.1 and 5.2, E does not intersect F. Clearly, each element of F is a limit element of E. Thus, $P(E)$ is not an amalgamation point. If $X \in E$, there exists an S-neighborhood of X, R, whose closure is perfectly compact in itself; by Theorems 10.3 and 11.3 $S-\bar{R}$ is an S-neighborhood of each boundary element. Thus, X is not a limit element of F, and $P(F)$ is an amalgamation point.

E 14.2. The point $Q(F)$ of E 9.2 is an amalgamation point; it is the amalgamation of all ends of F.

The point of E 8.3 is an amalgamation point but is not a composition point. The point $P(E)$ of E 8.2 is not an amalgamation point; here F is a limit element of E but does not intersect E.

Theorem 14.1. *An S-portion P is an amalgamation point if and only if the following condition holds: If the composition point X is a limit point of the collection of composition points M and $P<M<P$, then X intersects M and P.*

It may be proved with the help of D 14.1 and Theorems 11.2, 8.1, and 9.1.

Theorem 14.2. *A composition point is an amalgamation point.*

Proof. Let E and F be non-intersecting point elements; by D 4.4 there

(18) Cf. D 12.1, and D 7.1.

exist $e \in E$ and $f \in F$ such that $e \cdot f$ is vacuous. Then $R = S - \bar{e} \supset f$. If the point element G belongs to an element of $P(E)$, it intersects itself and E (cf. D 8.1 and Theorem 7.1). If $g \in G$, $e \cdot g$ is nonvacuous, R does not contain g, and R is not an S-neighborhood of G. Thus, F is not a limit element of any element of $P(E)$. By D 14.1 $P(E)$ is an amalgamation point.

D 14.3. If M is a point set, G is a collection of open sets in S, and for $m \in M$ there exists $g \in G$ such that g is an S-neighborhood of m, then G is said to *cover* M (*relative to* S).

THEOREM 14.3. *Let M be an amalgamable collection of ideal points and let G cover M relative to S. Then there exists a finite subcollection of G such that the sum of its elements is an S-neighborhood of M.*

Note that the conclusion does *not* state that the finite subcollection covers M relative to S. The latter is the type of conclusion one has where the Borel-Lebesgue property holds. Theorem 14.3 is used in establishing the Borel-Lebesgue property for certain decompositions of an amalgamation point (cf. Theorems 14.4, 14.5, and 15.3).

Proof. Suppose that the conclusion does not hold for G. Let N be the aggregate of all atomic ideal points which are portions of elements of M. If R is an open set of S, let $N(R)$ denote the aggregate of all elements of N of which R is an S-neighborhood. Let H be the collection of all R's such that $N(R)$ is covered relative to S by a finite subcollection of G, say $G(R)$. Let the symbol L denote the complement in S of an element of H; let E be the aggregate of all L's, and let K be the product of all the L's.

It follows from Theorems 8.1 and 9.1 that if an ideal point intersects an element of either M or N, then it intersects an element of the other; by D 9.2, $M < N$ and $N < M$; by Theorem 11.6 each S-neighborhood of M is an S-neighborhood of N, and conversely. Since each element of N is a portion of an element of M, it follows by Theorem 11.6 that G covers N; if $R \in G$, then R covers $N(R)$ relative to S; thus $R \in H$, and $H \supset G$. Suppose that D_1 is the sum of the elements of a finite subcollection of H, say H_1, and D_1 is an S-neighborhood of M; then D_1 is an S-neighborhood of N; by Theorem 11.7 H_1 covers N relative to S. If $R_i \in H_1$, $G(R_i)$ covers $N(R_i)$ relative to S. Since N is the sum of the finite collection of $N(R_i)$'s, N is covered by the sum of the sets $G(R_i)$; this sum, G^*, is a finite subcollection of G. If D^* is the sum of the elements of G^*, D^* is an S-neighborhood of N and of M. This is contrary to our supposition concerning G. Thus, D_1 and H_1 do not exist. Thus, the hypothesis of our theorem holds for each of G and H, but the conclusion holds for neither.

Suppose that an element of E, L, were vacuous; then $S - L = S$ would be an element of H; then $N = N(S)$, and N is covered by a finite subcollection of G, $G(S)$; since $H \supset G$ we are involved in a contradiction (cf. the preceding paragraph). Thus, L does not exist. If R_1 and R_2 are elements

of H, it follows from Theorem 11.7 that $N(R_1+R_2)=N(R_1)+N(R_2)$. Since $S-(R_1+R_2)=(S-R_1)\cdot(S-R_2)$, it follows that the product of two elements of E is an element of E.

Let $P\in K$. Suppose that $L_1\in E$ and P is not a limit point of L_1. There exists an open set in S, D, which contains P but contains no element of L_1-P. Then $S-L_1\supset D-P$ and $N(S-L_1)\supset N(D-P)$. Thus, $D-P\in H$ and $N(D-P)$ is covered relative to S by a finite subcollection of G. Since $P\in K\cdot D$, $S-D$ is not an element of E, D is not an element of H, $N(D)$ is not covered by a finite subcollection of G, and $N(D)\neq N(D-P)$. By Theorem 11.4 if P were non-isolated, we should have $N(D)=N(D-P)$; thus, P is isolated. By Theorem 8.4 there exists one atomic ideal point, β, such that $P=a(\beta)$, and β is degenerate (cf. D 12.2 and Theorem 13.7). Thus, $N(D)$ is either $N(D-P)$ or $\beta+N(D-P)$; in either case it can be covered by a finite subcollection of G, and $D\in H$. Therefore, we are involved in a contradiction, and L_1 does not exist.

Let $P\in K$. For R an S-neighborhood of P and $L\in E$ let $F(R, L)=(L-P)\cdot\overline{R}_S$, and let F be the aggregate of all $F(R, L)$'s. Since P is a limit point of each L, the product of two R's is an R, and the product of two L's is an L, it follows that the elements of F satisfy conditions (2) and (1) of D 4.1; if they also satisfy (4) they satisfy (3), and F is a decomposition point element. Suppose that (4) is not satisfied, and that U is a definite S-neighborhood of P which contains no element of F. Let T be the aggregate of all sets, $(\overline{U}-U)\cdot F(R, L)$, where $F(R, L)\in F$, and $U\supset R$. By the definition of U no element of T is vacuous. Since the product of two elements of F is an element of F, the product of two elements of T is an element of T. Let $Y\in\overline{U}-U$; since S is a Hausdorff space, there exist in it mutually exclusive open sets, V and W, which contain Y and P. If $R=U\cdot W$ and $L\in E$, then $(\overline{U}-U)\cdot F(R,L)$ does not contain Y; thus, the product of the elements of T is vacuous; by D 4.1 T is a boundary element. Similarly, if K is vacuous, E is a boundary element. Let Σ denote any of the three F, T, or E which is a point element; we have shown that at least one of them satisfies this condition. Σ has the property that if $L\in E$ then L contains an element of Σ. Let β be the composition point of which Σ is an element.

Let $X\in N$ and let Y be a point element which is an element of X (cf. Theorem 8.2 and D 8.4). There exists $g_X\in G$ such that g_X is an S-neighborhood of X (for, G covers N relative to S). Since $H\supset G$, $S-g_X\in E$, $S-g_X$ contains an element of Σ, and g_X contains an element of Y; by D 4.4 Σ does not intersect Y; by Theorem 8.1 β does not intersect X. Since $N<M<N$ and M is amalgamable so is N. By Theorems 9.1 and 14.1 β is not a limit point of N. By Theorem 12.2 there exists an S-neighborhood of β, D, which is an S-neighborhood of no element of N. There exists $\Sigma_1\in\Sigma$ such that $D\supset\Sigma_1$. Since $N(D)$ is vacuous, $D\in H$, $S-D\in E$, and $S-D\supset\Sigma_2$, where $\Sigma_2\in\Sigma$. Since $\Sigma_1\cdot\Sigma_2$ is vacuous, we are involved in a contradiction.

THEOREM 14.4. *An amalgamable collection of atomic ideal points has the Borel-Lebesgue covering property and is perfectly compact in itself.*

By Theorems 14.3 and 11.7 the set has the covering property. It is perfectly compact in itself; cf. Theorem 12.1 and Kuratowski and Sierpiński, loc. cit., Bibliography.

THEOREM 14.5. *The aggregate of all atomic ideal points is amalgamable and is perfectly compact in itself. The amalgamation of this collection is regular, and it is the maximal ideal point.*

Proof. Let M be the collection. By Theorem 9.1 the summation of the elements of M, β exists. If P and Q are intersecting ideal points, by Theorem 9.2 $\alpha(P)\cdot\alpha(Q)$ is nonvacuous; then $M\supset\alpha(P)\cdot\alpha(Q)$; by Theorems 9.2 and 9.1 Q intersects β and M; by D 9.2 $P<M$. By Theorem 14.1 β is an amalgamation point. Clearly, $a(\beta)$ consists of all points of S (cf. D 12.2 and D 9.4). By D 13.1 β is regular.

THEOREM 14.6. *Let P be a real point, Q be an amalgamation point, and M be a collection of composition points such that $Q<M<Q$. The following conditions are equivalent*: (1) *P and Q do not intersect*; (2) *no element of M intersects P*; (3) *P and Q can be separated by S-neighborhoods.*

Proof. Let $X\in M$ and E be a point element which belongs to some element of X. Since $X<Q$, by Theorem 8.1 E belongs to an element of Q. By D 9.1 if X intersects P, then so does Q; thus condition (1) implies condition (2). By condition (2) and Theorem 11.5 X and P can be separated by a pair of their respective S-neighborhoods, $U(X)$ and $V(X)$. By Theorem 14.3 there exists a finite collection of the $U(X)$'s such that the sum of these, U, is an S-neighborhood of M. If V is the product of the corresponding $V(X)$'s, V is an S-neighborhood of P, and $U\cdot V$ is vacuous. By Theorem 11.6 U is an S-neighborhood of Q. Thus, (2) implies (3). By D 11.1, D 11.2, D 9.1 and D 4.5, (3) implies (1).

THEOREM 14.7. *If P and Q are amalgamation points, one of which is regular, the following are equivalent*: (1) *P does not intersect Q*; (2) *P intersects no element of $\alpha(Q)$*; (3) *P and Q are separated by S-neighborhoods.*

Proof. By D 11.1, D 11.2, D 7.1, and D 4.4, and Theorem 8.1, (3) implies (1). By Theorem 9.1, (1) and (2) are equivalent; to establish (3) we shall consider the case of Q being regular. Let $X\in\alpha(P)$, $Y\in\alpha(Q)$, and E and F be point elements such that $E\in X$ and $F\in Y$ (cf. Theorem 8.2). By Theorem 8.1 E does not intersect F; there exist $e(Y)\in E$ and $f\in F$ such that $f\cdot e(Y)$ is vacuous. Then by D 4.1 $D(Y)=S-\overline{e(Y)}_S\supset f$, and $D(Y)$ is an S-neighborhood of F and of Y. By Theorems 9.1 and 14.3 there exist $Y_1, Y_2, \cdots, Y_n$ such that $D=D(Y_1)+D(Y_2)+\cdots+D(Y_n)$ is an S-neighborhood of $\alpha(Q)$ and of Q (cf. Theorem 11.6). By D 13.1 there exists an S-neighborhood of Q, R, such that

$D+a(Q)\supset \bar{R}_S$. By Theorem 9.1 $Q<\alpha(Q)<Q$; the methods of the proof of Theorem 12.4 show that $a(Q)=a(\alpha(Q))$. If each element of E contained points of $a(Q)$, by D 4.1, D 11.1, and D 12.2 X would be a limit point of $\alpha(Q)$; by Theorem 14.1 X intersects $\alpha(Q)$ and Q. By (1) this is impossible, and there exists $e_0 \in E$ such that $a(Q)\cdot e_0$ is vacuous. By D 4.1 there exists $e\in E$ such that $e_0\cdot e(Y_1)\cdot e(Y_2)\cdots e(Y_n)\supset e$. Then $e\cdot(D+a(Q))$ is vacuous; $S-\bar{R}_S$ contains e and is an S-neighborhood of E and X; X and Q can be separated by S-neighborhoods. The conclusion for (3) follows by methods used for Theorem 14.6.

THEOREM 14.8. *A regular amalgamation point which is a portion of either a decomposition point or of a collection of boundary points is a composition point.*

Proof. Let P be the amalgamation point and Q be a decomposition point such that $P<Q$. Let $E\in P$; let F and G be point elements such that $F\in Q$ and G belongs to E; let $F(P)$ and $Q(P)$ be defined as in Theorem 13.7. By Theorem 7.2 and D 8.2 $G<E<F$. By D 7.1 G intersects F. By Theorems 5.1 and 12.4, and D 12.2, $a(Q)=a(F)=a(G)=a(E)=a(P)$. If R is an S-neighborhood of $a(P)$, it is an S-neighborhood of F, Q, and P (cf. D 4.1, D 11.1, and Theorem 11.6). Since P is regular, it has an S-neighborhood U such that $R+a(P)\supset \bar{U}_S$. Then $\bar{U}_S$ or $\bar{U}_S-a(P)$ is an element of $F(P)$. By D 4.1 $F(P)$ is a point element, and $a(P)=a(F(P))$.

Consider the second hypothesis for P. Let $X\in S$, $R=S-X$; by an argument similar to that of the preceding paragraph, it follows that $a(P)$ is vacuous, and $F(P)$ is a boundary element.

Since each element of $F(P)$ contains an S-neighborhood of P, each element of $F(P)$ contains elements of each point element which belongs to E. Thus, $E<F(P)$, and $P<Q(P)$; cf. Theorem 7.1 and D 8.2. Suppose that *not* $Q(P)<P$; by Theorem 9.2 there exists $Y\in\alpha(Q(P))-\alpha(P)$, and Y does not intersect P. By Theorem 14.7 P and Y can be separated by S-neighborhoods U and V. Since $\bar{U}$ or $\bar{U}-a(P)$ is an element of $F(P)$, and $F(P)\in Q(P)$, $F(P)$ intersects no point element of which V is an S-neighborhood. By Theorem 8.1 Y and $Q(P)$ do not intersect. Thus, the supposition that "*not* $Q(P)<P$" involves a contradiction. Since, also, $P<Q(P)$, $P=Q(P)$.

THEOREM 14.9. *If D is an S-neighborhood of the regular amalgamation point P, then the set $(S-D)\cdot a(P)$ is finite or vacuous.*

Proof. Suppose that $M=(S-D)\cdot a(P)$ is infinite. By Theorems 10.3, 10.4, and 12.3 there exists a composition point β which is an end and a limit point of M; since $S-D\supset M$, β is an end of $S-D$ (cf. D 10.2). By Theorems 10.2 and 11.3 P and β do not intersect. By Theorem 14.7 P and β can be separated by S-neighborhoods U and V. Since β is a limit point of M, V contains points of M and of $a(P)$; by Theorem 12.4 $\bar{U}\supset a(P)$. Thus, the supposition that M is infinite involves a contradiction.

THEOREM 14.10. (1) *A perfectly compact set of regular amalgamation points is amalgamable.* (2) *An amalgamation of a collection of regular amalgamation points is regular.*

Proof. By Theorem 9.1, if N is the set of (1), then N has a summation P; $N<P<\alpha(P)<N$. If the composition point X intersects no element of $\alpha(P)$, by D 9.2 X intersects no element of N. By Theorem 14.7 each element of N can be separated from X by S-neighborhoods. There exists a finite collection H of open sets which covers N relative to S, and an S-neighborhood of X, U, such that U has no point in common with an element of H (cf. the proof of Theorem IX, p. 89, Alexandroff and Hopf). Let W be the sum of the elements of H; it is an S-neighborhood of N and of $\alpha(P)$ (cf. Theorem 11.6). Thus, X is not a limit point of $\alpha(P)$. By Theorem 14.1 P is an amalgamation point.

Let the elements of N be regular, and let D be an S-neighborhood of P. If $X\in N$, D is an S-neighborhood of X and $a(P)\supset a(X)$ (cf. Theorems 11.6 and 12.4). By our hypothesis there exists an S-neighborhood of X, $U(X)$, such that $D+a(X)\supset \overline{U(X)}$. By Theorems 14.3 and 11.6 there exists a finite collection of the $U(X)$'s, whose sum U is an S-neighborhood of N and of P. Then $D+a(P)\supset\overline{U}$.

THEOREM 14.11. (1) *If a collection H of amalgamation points has a lower bound, its greatest lower bound is an amalgamation point*; (2) *if the elements of H are regular and one of them is the amalgamation of a set of regular composition points, then their greatest lower bound is regular.*

Theorems 14.5, 14.10, and 14.11 are applied in Chapter III in questions concerned with the existence and the properties of multiplicative systems and complete lattices of regular amalgamation points. Examples may be given to show that the conclusion of (2) does not hold if the second part of the hypothesis be omitted.

Proof. Let F be the product of all the $\alpha(h)$'s, where $h\in H$. If P is the summation of the elements of F, by Theorem 9.2 P is the greatest lower bound of the h's. Let the composition point Q be a limit point of F which does not intersect P. Suppose that no element of $\alpha(Q)$ is a limit point of F. By Theorem 12.2, if $\beta\in\alpha(Q)$, there exists an S-neighborhood of β, $D(\beta)$, which is not an S-neighborhood of any element of F. By Theorems 14.2, 14.3, and 11.6 there exists an S-neighborhood of Q, D, which is the sum of a finite collection of the $D(\beta)$'s. By Theorem 11.7 D is not an S-neighborhood of any element of F. Thus, we are involved in a contradiction, and there exists $\beta\in\alpha(Q)$ such that β is a limit point of F. If $h\in H$, $\alpha(h)\supset F$, and β is a limit point of $\alpha(h)$; by Theorem 14.1 β intersects an element of $\alpha(h)$. Since β is atomic, $\beta\in\alpha(h)$. Thus, $\beta\in F$; by Theorem 8.1 P and Q intersect. By Theorem 14.1 P is an amalgamation point.

We shall first establish (2) for the case where the number of elements of H is the positive integer n. Clearly (2) holds for $n=1$. Suppose that (2) holds for each collection with fewer than n elements. Let A be an element of H which is the amalgamation of a collection T of regular amalgamation points; Let $B\in H-A$, and $H'=H-B$. By the definition of n, the elements of H' have a greatest lower bound C, which is a regular amalgamation point. Let Z be the greatest lower bound of B and C. By Theorem 9.2 Z is the greatest lower bound of the elements of H; by part (1) Z is an amalgamation point. Let D be an S-neighborhood of Z. If X is an amalgamation point, let $D(X)$ denote the aggregate of all elements of $\alpha(X)$ of which D is *not* an S-neighborhood. By Theorem 14.4 $\alpha(X)$ is perfectly compact in itself. Since $D(X)$ is closed relative to $\alpha(X)$, it is perfectly compact in itself (cf. Fréchet, (I), pp. 229–230). By Theorem 11.6 D is an S-neighborhood of $\alpha(Z)$; by Theorem 9.2 $\alpha(Z)=\alpha(B)\cdot\alpha(C)$; it follows that $D(B)\cdot D(C)$ is vacuous. If $X\in D(B)$, by Theorem 8.1 X does not intersect C. By the definition of n, C is regular; by Theorem 14.7 X and C can be separated by S-neighborhoods. By Theorem 11.6 each element of $D(B)$ can be separated from each element of $D(C)$ by S-neighborhoods. Since each of $D(B)$ and $D(C)$ is perfectly compact in itself, they can be separated by S-neighborhoods D_1 and D_2 (apply an argument similar to that for the proof of Theorem IX, p. 89, Alexandroff and Hopf). Then $D+D_1$ and $D+D_2$ are S-neighborhoods, respectively, of $\alpha(B)$ and of B, and of $\alpha(C)$ and of C, respectively. Since B and C are regular, there exist S-neighborhoods R_1 and R_2 of B and C, respectively, such that $E_1=D+D_1+a(B)\supset\overline{R}_1$, and $E_2=D+D_2+a(C)\supset\overline{R}_2$. Since $Z<B$ and $Z<C$, by Theorem 12.1 $R=R_1\cdot R_2$ is an S-neighborhood of Z. Let $M_1=D+D_1$, and $N_1=(S-M_1)\cdot a(B)$; thus, $M_1+N_1=E_1$. Let $M_2=D+D_2$, $N_2=(S-M_2)\cdot a(C)$, and $K=M_1\cdot N_2+M_2\cdot N_1+N_1\cdot N_2$. By Theorem 14.9 N_1, N_2, and K are finite or vacuous. Since $D_1\cdot D_2$ is vacuous, $M_1\cdot M_2=D$, and $E_1\cdot E_2=(M_1+N_1)\cdot(M_2+N_2)=D+K$. Thus, $D+K\supset\overline{R}$. By Theorem 12.4 $\overline{R}\supset a(Z)$ and $a(B)\cdot a(C)\supset a(Z)$. Let $L=K-K\cdot a(Z)$; then L is finite or vacuous. Suppose that there exists $X\in\alpha(Z)$ and $Y\in L$ such that X intersects Y. Since $X<Z<A$, $X\in\alpha(A)$. Since $A<T$, by D 9.2 X intersects an element of T, say X_1. Let the point elements E and E_1 be elements of X and of X_1, respectively. By Theorem 8.1 E and E_1 intersect; by Theorem 5.1 $a(E)=a(E_1)$; by D 12.2 $a(X)=a(X_1)$. Since $X<Z$, by Theorem 12.4 $a(Z)\supset a(X)$; thus, $a(X_1)\neq Y$. By D 4.1 there exists $e_1\in E_1$ such that e_1 does not contain Y. Then $S-Y$ is an S-neighborhood of E_1 and of X_1. Since X_1 is regular, there exists an S-neighborhood of X_1, U, such that $S-Y\supset\overline{U}$; then $S-\overline{U}$ is an S-neighborhood of Y. Since X is atomic and intersects X_1, $X<X_1$; then U is an S-neighborhood of X. By Theorem 11.5 X does not intersect Y; thus the supposition that X and Y exist involves a contradiction. By Theorem 14.6 Z and L can be separated by a pair of their respective S-neighborhoods, U and V. Then $W=R\cdot U$ is an S-neighborhood of Z, $\overline{R}\supset\overline{W}$, and $\overline{W}\cdot L$ is vacu-

ous; since $D+K\supset\bar{R}$, $D+a(Z)\supset D+(K-L)\supset\bar{W}$. Thus, Z is regular. Thus, by induction we establish the conclusion of (2) for any case where H is finite.

Consider the infinite case. Let Z be the lower bound of the elements of H; let A, D, and $D(X)$ be defined as in the preceding paragraph. By Theorem 9.2 $\alpha(Z)$ is the product of the sets $\alpha(h)$, where $h\in H$; by Theorems 9.1 and 11.6 D is an S-neighborhood of $\alpha(Z)$; hence, the sets $D(h)$ have a vacuous product. We showed in the preceding paragraph that each of the $D(h)$'s is perfectly compact in itself; it follows that there exists a finite subcollection of H, $H'=(h_1, h_2, \cdots, h_n)$, such that the product $\lambda=D(h_1)\cdot D(h_2)\cdots D(h_n)$ is vacuous (cf. Fréchet, (I), p. 231). We shall suppose that $A\in H'$; this is permissible, for if the product λ is vacuous, the product of all the elements of λ by $D(A)$ will also be vacuous. Let J be the product $\alpha(h_1)\cdot\alpha(h_2)\cdot\alpha(h_3)\cdots\alpha(h_n)$. Let $B=C=P(J)$; by the results of the preceding paragraph B is the greatest lower bound of the elements of H', and it is a regular amalgamation point. Clearly D is an S-neighborhood of each element of J and of J; by Theorem 11.6 D is an S-neighborhood of B. Let R be an S-neighborhood of B such that $D+a(B)\supset\bar{R}$; let $K=(S-D)\cdot a(B)$ and $L=K-K\cdot a(Z)$. Then $D+K\supset\bar{R}$. By following the argument of the later portion of the preceding paragraph, we see that Z is a regular amalgamation point.

Theorem 14.12. (1) *A Hausdorff space is normal if and only if all its atomic boundary points are regular*; (2) *it is semi-completely normal if and only if all its atomic ideal points are regular.*

Proof. By Theorems 13.2 and 13.3 the conditions are necessary. By Theorem 14.2 each composition point is an amalgamation point. If the conditions hold it follows from Theorems 14.10, 14.2, and 9.1 that each of the composition points involved in the conditions of Theorems 13.2 and 13.3 is regular. Thus, the conditions are sufficient.

Theorem 14.13. *All the amalgamation points of a semi-completely normal space are regular, and each of them can be decomposed into a collection of regular atomic composition points.*

This follows from Theorems 9.1, 14.12 and 14.10.

Examples. E 14.3. Urysohn([16]) has constructed a countable space R such that if X and Y are points of R and U and V are open sets containing X and Y, respectively, then $\bar{U}\cdot\bar{V}$ is nonvacuous; because of this condition no point of R is isolated. Let P be an amalgamation point which is distinct from the maximal amalgamation point of R; cf. D 9.4 and Theorem 14.5. Suppose that P is regular. By D 9.2 and Theorem 8.1 there exists a composition point Q which does not intersect P. By Theorems 14.2 and 14.7 P and Q can be separated by a pair of their respective R-neighborhoods, U and V. There exists an

([16]) Loc. cit., pp. 274–283.

R-neighborhood of P, W, such that $U+a(P)\supset\overline{W}$. By Theorem 14.9 there exist at most a finite number of points in $(R-U)\cdot a(P)$, say $P_1, P_2, \cdots, P_n$. Since no point of R is isolated it follows from Theorems 12.3, 11.3, and 11.6 that each of P and Q is a limit point of R. Thus, there exist $X\in V$ and $Y\in W$. For each i let δ_i and λ_i be mutually exclusive open sets which contain P_i and X respectively. Let $\lambda=V\cdot\lambda_1\cdot\lambda_2\cdots\lambda_n$. Then $\overline{V}\supset\overline{\lambda}$, $U\cdot\overline{V}$ is vacuous, $a(P)\cdot\overline{\lambda}$ and $\overline{W}\cdot\overline{\lambda}$ are vacuous; we are involved in a contradicion. *To summarize*: the maximal ideal point is a regular amalgamation point of R and is the only one; by Theorem 14.2 no composition point of R is regular.

15. **The upper semi-continuity and the perfect compactness of decompositions.** First we shall recall definitions by Moore and by Alexandroff(17), and then we shall introduce extensions of these. (I) Let T be a space H Fréchet, K be a point set of T, and M be a collection of mutually exclusive subsets of K whose sum is K; M is called a *decomposition* (*Zerlegung*) of K (cf. D 9.3). (II) A collection M of mutually exclusive closed point sets of T is said to be *upper semi-continuous* (*in* T) provided that if $P\in M$ and D is an open set of T which contains P, there exists in T an open set R which contains P such that if an element of M has a point in common with R then that element is a subset of D(18). (III) If $T\supset K$ and M is a decomposition of K, then $T(M)$, the (*weak*) *space of the decomposition*, is defined as follows: (1) its points are the elements of M; (2) if U is an open point set in T, let $U(M)$ be the aggregate of all elements of M which are subsets of U; the $U(M)$'s are the *neighborhoods* for $T(M)$(19) (cf. Theorem 12.1). (IV) A continuous mapping of the space X on the space Y is said to be *closed* if the image of each closed point set in X is a closed point set in Y(20).

Examples. E 15.1. Let X be the subspace of the plane whose points are those of a circle and its interior. Let δ be a definite diameter of the circle; Let Y_1 be the decomposition of X which consists of δ and all chords of the circle which are parallel to δ; let Z_1 be the decomposition whose elements are the points of δ, and the chords that are parallel to δ. Let Y and Z be the spaces of the decompositions Y_1 and Z_1 respectively. Let the relations $y=f(x)$ and $z=g(x)$ mean, respectively, that $x\in y$ and $x\in z$. The collection Y_1 is upper semi-continuous in X, but Z_1 is not. The spaces X and Y are perfectly compact, but Z is not. The mapping $y=f(x)$ of X on Y is closed. The mapping

(17) Cf. Moore, (III), and Alexandroff, (I). For treatments of the theory see Moore, (I), chap. 5, and Alexandroff and Hopf, pp. 61–70 and 95–98.

(18) Moore, loc. cit.; Alexandroff and Hopf call M a *continuous decomposition* (*stetige Zerlegung*) of the sum of the elements of M (loc. cit., p. 67).

(19) Alexandroff and Hopf, p. 66, call $T(M)$ *der schwache Zerlegungsraum der Zerlegung M.* If M is upper semi-continuous, $T(M)$ is their *Zerlegungsraum* (loc. cit., pp. 63 and 67). We deal mainly with the latter.

(20) Alexandroff and Hopf, p. 95; for information and definitions having to do with mappings (*Abbildungen*) see pages 51, 52; in V, page 54, they show that a continuous mapping need not be closed.

ous; since $D+K\supset\overline{R}$, $D+a(Z)\supset D+(K-L)\supset\overline{W}$. Thus, Z is regular. Thus, by induction we establish the conclusion of (2) for any case where H is finite.

Consider the infinite case. Let Z be the lower bound of the elements of H; let A, D, and $D(X)$ be defined as in the preceding paragraph. By Theorem 9.2 $\alpha(Z)$ is the product of the sets $\alpha(h)$, where $h\in H$; by Theorems 9.1 and 11.6 D is an S-neighborhood of $\alpha(Z)$; hence, the sets $D(h)$ have a vacuous product. We showed in the preceding paragraph that each of the $D(h)$'s is perfectly compact in itself; it follows that there exists a finite subcollection of H, $H'=(h_1, h_2, \cdots, h_n)$, such that the product $\lambda=D(h_1)\cdot D(h_2)\cdots D(h_n)$ is vacuous (cf. Fréchet, (I), p. 231). We shall suppose that $A\in H'$; this is permissible, for if the product λ is vacuous, the product of all the elements of λ by $D(A)$ will also be vacuous. Let J be the product $\alpha(h_1)\cdot\alpha(h_2)\cdot\alpha(h_3)\cdots\alpha(h_n)$. Let $B=C=P(J)$; by the results of the preceding paragraph B is the greatest lower bound of the elements of H', and it is a regular amalgamation point. Clearly D is an S-neighborhood of each element of J and of J; by Theorem 11.6 D is an S-neighborhood of B. Let R be an S-neighborhood of B such that $D+a(B)\supset\overline{R}$; let $K=(S-D)\cdot a(B)$ and $L=K-K\cdot a(Z)$. Then $D+K\supset\overline{R}$. By following the argument of the later portion of the preceding paragraph, we see that Z is a regular amalgamation point.

THEOREM 14.12. (1) *A Hausdorff space is normal if and only if all its atomic boundary points are regular*; (2) *it is semi-completely normal if and only if all its atomic ideal points are regular.*

Proof. By Theorems 13.2 and 13.3 the conditions are necessary. By Theorem 14.2 each composition point is an amalgamation point. If the conditions hold it follows from Theorems 14.10, 14.2, and 9.1 that each of the composition points involved in the conditions of Theorems 13.2 and 13.3 is regular. Thus, the conditions are sufficient.

THEOREM 14.13. *All the amalgamation points of a semi-completely normal space are regular, and each of them can be decomposed into a collection of regular atomic composition points.*

This follows from Theorems 9.1, 14.12 and 14.10.

EXAMPLES. E 14.3. Urysohn[16] has constructed a countable space R such that if X and Y are points of R and U and V are open sets containing X and Y, respectively, then $\overline{U}\cdot\overline{V}$ is nonvacuous; because of this condition no point of R is isolated. Let P be an amalgamation point which is distinct from the maximal amalgamation point of R; cf. D 9.4 and Theorem 14.5. Suppose that P is regular. By D 9.2 and Theorem 8.1 there exists a composition point Q which does not intersect P. By Theorems 14.2 and 14.7 P and Q can be separated by a pair of their respective R-neighborhoods, U and V. There exists an

[16] Loc. cit., pp. 274–283.

R-neighborhood of P, W, such that $U+a(P)\supset\overline{W}$. By Theorem 14.9 there exist at most a finite number of points in $(R-U)\cdot a(P)$, say $P_1, P_2, \cdots, P_n$. Since no point of R is isolated it follows from Theorems 12.3, 11.3, and 11.6 that each of P and Q is a limit point of R. Thus, there exist $X\in V$ and $Y\in W$. For each i let δ_i and λ_i be mutually exclusive open sets which contain P_i and X respectively. Let $\lambda=V\cdot\lambda_1\cdot\lambda_2\cdots\lambda_n$. Then $\overline{V}\supset\bar{\lambda}$, $U\cdot\overline{V}$ is vacuous, $a(P)\cdot\bar{\lambda}$ and $\overline{W}\cdot\bar{\lambda}$ are vacuous; we are involved in a contradicion. *To summarize*: the maximal ideal point is a regular amalgamation point of R and is the only one; by Theorem 14.2 no composition point of R is regular.

15. **The upper semi-continuity and the perfect compactness of decompositions.** First we shall recall definitions by Moore and by Alexandroff([17]), and then we shall introduce extensions of these. (I) Let T be a space H Fréchet, K be a point set of T, and M be a collection of mutually exclusive subsets of K whose sum is K; M is called a *decomposition* (*Zerlegung*) of K (cf. D 9.3). (II) A collection M of mutually exclusive closed point sets of T is said to be *upper semi-continuous* (*in* T) provided that if $P\in M$ and D is an open set of T which contains P, there exists in T an open set R which contains P such that if an element of M has a point in common with R then that element is a subset of D([18]). (III) If $T\supset K$ and M is a decomposition of K, then $T(M)$, the (*weak*) *space of the decomposition*, is defined as follows: (1) its points are the elements of M; (2) if U is an open point set in T, let $U(M)$ be the aggregate of all elements of M which are subsets of U; the $U(M)$'s are the *neighborhoods* for $T(M)$([19]) (cf. Theorem 12.1). (IV) A continuous mapping of the space X on the space Y is said to be *closed* if the image of each closed point set in X is a closed point set in Y([20]).

EXAMPLES. E 15.1. Let X be the subspace of the plane whose points are those of a circle and its interior. Let δ be a definite diameter of the circle; Let Y_1 be the decomposition of X which consists of δ and all chords of the circle which are parallel to δ; let Z_1 be the decomposition whose elements are the points of δ, and the chords that are parallel to δ. Let Y and Z be the spaces of the decompositions Y_1 and Z_1 respectively. Let the relations $y=f(x)$ and $z=g(x)$ mean, respectively, that $x\in y$ and $x\in z$. The collection Y_1 is upper semi-continuous in X, but Z_1 is not. The spaces X and Y are perfectly compact, but Z is not. The mapping $y=f(x)$ of X on Y is closed. The mapping

([17]) Cf. Moore, (III), and Alexandroff, (I). For treatments of the theory see Moore, (I), chap. 5, and Alexandroff and Hopf, pp. 61–70 and 95–98.

([18]) Moore, loc. cit.; Alexandroff and Hopf call M a *continuous decomposition* (*stetige Zerlegung*) of the sum of the elements of M (loc. cit., p. 67).

([19]) Alexandroff and Hopf, p. 66, call $T(M)$ *der schwache Zerlegungsraum der Zerlegung* M. If M is upper semi-continuous, $T(M)$ is their *Zerlegungsraum* (loc. cit., pp. 63 and 67). We deal mainly with the latter.

([20]) Alexandroff and Hopf, p. 95; for information and definitions having to do with mappings (*Abbildungen*) see pages 51, 52; in V, page 54, they show that a continuous mapping need not be closed.

$z=g(x)$ of X on Z is not continuous. The mapping $y=f(g^{-1}(z))$ of Z on Y is continuous but is not closed. The example both serves to illustrate the results of Theorem 15.1 and to give *Gegenbeispiele*. If we amalgamate the elements of Y_1 and Z_1, we obtain collections of amalgamation points which may be used to illustrate Theorem 15.4.

The extensive results of Alexandroff and Hopf in their Theorem VIII, p. 98, may be amplified by the following theorem. The mapping of their hypothesis is continuous rather than closed. They give no result analogous to that of the Conditions (I) and (III) implying (II); the transformation $y=f(g^{-1}(z))$ of the preceding example illustrates the fact that the stronger condition, *closedness*, is essential for such a result.

THEOREM 15.1. *Let X and Y be Hausdorff spaces, and let $y=\alpha(x)$ define a continuous mapping of X into Y; then any two of the following conditions imply the third*: (I) *Y is perfectly compact*; (II) *X is perfectly compact*; (III) *the mapping $y=\alpha(x)$ is closed; for $y\in Y$ the set $\alpha^{-1}(y)$ is perfectly compact in itself.*

Proof. By Alexandroff and Hopf, Theorem VIII, p. 98, Conditions (II) and (III) imply (I). By this theorem and their Theorems II on page 95, (I) on page 53, and IV on page 86 Conditions (I) and (II) imply (III). Let H be a collection of open sets which covers X. If $y\in Y$ let $H(y)$ be a finite subcollection of H which covers $\alpha^{-1}(y)$ (cf. Condition (III) and Kuratowski and Sierpiński, loc. cit.). Let $D(y)$ be the sum of the elements of $H(y)$; then $X-D(Y)$ is closed in X. By (III) $\alpha(X-D(y))$ is closed in Y; and $\delta(y)=Y-\alpha(X-D(y))$ is open in Y and contains y. By (I) there exists a finite subcollection of the $\delta(y)$'s which covers Y, say $\delta(y_1)$, $\delta(y_2)$, $\cdots$, $\delta(y_n)$. Let $H^*=H(y_1)+H(y_2)+\cdots+H(y_n)$. If $x_0\in X$, there exists j such that $\alpha(x_0)\in\delta(y_j)$; if x_0 were not covered by $D(y_j)$, it would belong to $X-D(y_j)$; then $\alpha(x_0)\in\alpha(X-D(y_j))=Y-\delta(y_j)$, and we are involved in a contradiction. Thus x_0 is covered by an element of $H(y_j)$, and X is covered by H^*. Thus, (I) and (III) imply (II).

D 15.1. A collection M of points is said to be *upper semi-continuous* (*relative to S*) provided that if $P\in M$ and D is an S-neighborhood of P, there exists an S-neighborhood of P, R, such that if Q is an element of M and R is an S-neighborhood of a portion of Q, then D is an S-neighborhood of Q (cf. Definition (II) above).

EXAMPLES. E 15.2. Let M be a decomposition of the perfectly compact Hausdorff space S into closed sets. Let N be the collection of those ideal points which are amalgamations of elements of M. Then N is upper semi-continuous relative to S if and only if M is upper semi-continuous in S. Examples of both possibilities are given by Alexandroff and Hopf, p. 67. See also E 15.1.

E 15.3. A collection of real points, or of atomic ideal points, is upper semi-continuous relative to S.

If, in Theorem 15.1, we amalgamate the elements of the sets $\alpha^{-1}(y)$, for

$y \in Y$, we obtain amalgamation points for which Theorem 15.4 applies. This is an instance of the analogy we have noted before between the properties of perfectly compact sets and of amalgamation points. Similarly, the results of Theorem 15.3 resemble some of those of Alexandroff and Hopf. These analogies give indications of some rather trivial applications of our theory. In general, our procedure is to define mappings in terms of decompositions or of amalgamations; these processes involve mainly order relations. This procedure must be carried out in such a manner that it gives conditions which assure the continuity of the mappings and permit the application of the results of Alexandroff and Hopf. Such conditions are given in Theorems 15.4, 15.5, and 15.6. Thus, we have on the one hand the processes of amalgamation of points and of continuous mappings, and on the other those of decomposition of points and of the inverse of a continuous mapping. We may think of the latter as giving a kind of representation theory([21]) for our system of points, in which the representation of a point is a decomposition of that point into an upper semi-continuous collection of amalgamation points. Thus, this representation theory involves not only order, but also continuity.

THEOREM 15.2. *If N is an upper semi-continuous collection of amalgamation points, no two of them intersect.*

Suppose that X and Y are intersecting elements of N. By D 15.1 X and Y have the same S-neighborhoods. By Theorem 9.2 if not $X < Y$ then there exists an element β of $\alpha(X) - \alpha(X) \cdot \alpha(Y)$. By Theorem 12.2 if $\lambda \in \alpha(Y)$ there exists an S-neighborhood of λ, $\delta(\lambda)$, which is not an S-neighborhood of β. By Theorem 14.3 there exists a finite collection of the $\delta(\lambda)$'s, whose sum D is an S-neighborhood of Y. By Theorem 11.7 D is not an S-neighborhood of β. Since $\beta < X$ and D is an S-neighborhood of X, we are involved in a contradiction. Thus, $X < Y$. Similarly, $Y < X$. Hence, $Y = X$.

THEOREM 15.3. *Let Y be a decomposition*([22]) $[x]$ *of the ideal point X into regular amalgamation points. The following conditions are equivalent:* (1) *Y is perfectly compact in itself*; (2) *X is an amalgamation point and Y is upper semi-continuous relative to S*; (3) *the space*([23]) $S(Y)$ *of the decomposition Y is a perfectly compact Hausdorff space.*

The following is a point set analogue of this theorem.

THEOREM 15.3A. *If Y is a decomposition of the Hausdorff space X into a collection of sets which are perfectly compact in themselves, the following conditions are equivalent:* (1) *The space $X(Y)$ of the decomposition Y is perfectly compact*; (2) *X is perfectly compact and Y is upper semi-continuous in X.*

([21]) Cf. Birkhoff, p. 76, and Stone, loc. cit.
([22]) Cf. D 9.3.
([23]) Cf. Theorem 12.1.

Consider 15.3A. By Alexandroff and Hopf, p. 98, Theorem VIII, (2) implies (1). Let (1) hold and $y=\alpha(x)$ mean that $x\in y\in Y$ $(X\supset y)$. If U is an open set in X, $U(Y)$ means the set of all elements of Y that are subsets of U; and $U(Y)$ is an open set in $X(Y)$ (cf. Definition (III)). Let $y\in Y$, and R be an open set in X containing y. If $z\in Y-R(Y)$, there exist in X mutually exclusive open sets, V_z and W_z, which contain, respectively, y and z (cf. proof, Theorem IX, p. 89, Alexandroff and Hopf). Since $Y-R(Y)$ is a closed set in the perfectly compact space $X(Y)$, it has the Borel-Lebesgue property; there exist $W_{z_1}(Y)$, $W_{z_2}(Y)$, $\cdots$, $W_{z_n}(Y)$ which cover $Y-R(Y)$ in $X(Y)$. Let $V=V_{z_1}\cdot V_{z_2}\cdots V_{z_n}$ and $W=W_{z_1}+W_{z_2}+\cdots+W_{z_n}$. Then V and W are mutually exclusive open sets in X, $V\supset y$, and W contains all elements of $Y-R(Y)$. Since $V\cdot W$ is vacuous, any element of Y that has a point in common with V is a subset of R. Thus, Y is upper semi-continuous in X; and the mapping $y=\alpha(x)$ of X on $X(Y)$ is continuous (cf. Alexandroff and Hopf, p. 67). Let F be a closed point set in X, and $D=X-F$. Then $D(Y)$ is open in $X(Y)$. Since $D(Y)=X(Y)-\alpha(F)$, $\alpha(F)$ is closed in $X(Y)$, and the mapping is closed. By Theorem 15.1 Condition (1) implies (2).

Proof of Theorem 15.3. Let (2) hold and G be a collection of open sets of S that covers Y relative to S. If $x\in Y$, let $D\in G$ such that D is an S-neighborhood of x. Let $R(x, D)$ be an S-neighborhood of x such that if it is an S-neighborhood of a portion of an element of Y, then D is an S-neighborhood of that element. By Theorems 14.4, 11.6, and 9.1 there exists a finite collection of the R's which covers the aggregate of atomic portions of X; say $R(x_1, D_1)$, $R(x_2, D_2)$, $\cdots$, $R(x_n, D_n)$. Let G^* be the collection $D_1, D_2, \cdots, D_n$. If $x\in Y$ and β is an atomic portion of x, $\beta<x<X$, and there exists j such that $R(x_j, D_j)$ is an S-neighborhood of β. Then D_j is an S-neighborhood of x, and G^* covers Y relative to S. Since Y has the Borel-Lebesgue covering property, it is perfectly compact in itself (cf. Theorem 12.1 and Kuratowski and Sierpiński, loc. cit.). Thus (2) implies (1).

Conversely, let Y be perfectly compact in itself. By Theorem 14.10 X is an amalgamation point. Let $x\in Y$, D be an S-neighborhood of x, and L be the set of all elements of Y of which D is not an S-neighborhood. By D 12.1 no point of $Y-L$ is a limit point of L, and L is a closed point set relative to Y. Since Y is perfectly compact in itself, so is L. By Theorem 14.10 there exists a regular amalgamation point β, such that $\beta<L<\beta$. By D 9.3 x intersects no element of L; and thus it does not intersect β. By Theorem 14.7 x and β can be separated by a pair of their respective S-neighborhoods, U and W; by Theorem 11.6 W is an S-neighborhood of L. Let x_1 be an element of Y such that U is an S-neighborhood of a portion λ of x_1. If x_1 were an element of L, W would be an S-neighborhood of x_1 and of λ; since $U\cdot W$ is vacuous, this involves a contradiction. Since $x_1\in Y-L$, D is an S-neighborhood of x_1. Thus, Y is upper semi-continuous, and (1) implies (2).

If x_1 and x_2 are two elements of Y, by Theorem 14.7 and D 9.3 they can

be separated by S-neighborhoods. By Theorem 12.1 condition (1) implies (3). The converse is obvious.

THEOREM 15.4. *Given that P is an amalgamation point and Y and X are decompositions of P into regular amalgamation points such that if $y \in Y$ there exists a subcollection $\alpha^{-1}(y)$ of X which is a decomposition of y; then any two of the following conditions imply the third*: (1) *Y is perfectly compact in itself*; (2) *X is perfectly compact in itself*; (3) *the mapping $y=\alpha(x)$ of X into Y is a closed mapping*; *the sets $\alpha^{-1}(y)$ for $y \in Y$, are perfectly compact in themselves.*

By Theorem 15.3 the conditions (2) and (1) are equivalent, respectively, to the upper semi-continuity of X and of Y. Cf. E 15.1.

The close resemblance of Theorems 15.1 and 15.4 fails to extend to the explicit assumption of the continuity of the mapping $y=\alpha(x)$, which is necessary for the former. The inherent potentialities for the continuity of $y=\alpha(x)$ for Theorem 15.4 are an example of the properties of this mapping which follow from the way it is defined: For $y \in Y$, $y<\alpha^{-1}(y)<y$; $X \supset \alpha^{-1}(y)$; and if $x \in \alpha^{-1}(y)$, then $x<y$ (cf. D 9.3). These order relations, which involve X, Y, and the mapping, may be regarded as a key to our theory, and are a foundation for the applications we make in following sections. The transformation expresses the fact that the elements of y are obtained by amalgamations of elements of X; conversely, the elements of X are the results of decompositions of the y's. For topological applications we need conditions of continuity as well as those for order; simultaneous conditions for the required continuity and order are given in Theorems 15.3 to 15.6.

Proof. By Theorem 15.3 if condition (1) or (2) holds, then Y or X, respectively, is upper semi-continuous relative to S. By (1) and (2) the mapping $y=\alpha(x)$ is continuous from X to Y (cf. argument on page 67, and Theorem IV, p. 53, Alexandroff and Hopf). The conclusion follows from Theorem 15.1.

THEOREM 15.5. *Adopt the notation of Theorem* 15.4, *and let X be upper semi-continuous*(24) *relative to S*; *then* (A) *Y is upper semi-continuous relative to S if and only if the aggregate $Y'=[\alpha^{-1}(y)]$, where $y \in Y$, is an upper semi-continuous collection of point sets in X*(25). (B) *Let $Y''=X(Y')$ be the space*(26) *of the decomposition of X, Y', and let $y''=\alpha''(x)$ mean that $x \in y'' \in Y''$; then the mapping $y=\alpha(x)$ of X on Y is closed if and only if the mapping $y''=\alpha''(x)$ of X on Y'' is closed.* (C) *If the conditions in either* (A) *or* (B) *are satisfied, then those in each of Theorem* 15.4, (A) *and* (B), *are satisfied, and the mappings $y=\alpha(x)$ and $y''=\alpha''(x)$ are equivalent mappings of X*(27).

(24) Or, X is perfectly compact in itself (cf. Theorem 15.3).

(25) Cf. definition (II). Here X is regarded as a space (cf. Theorem 12.1).

(26) Cf. Definition (III); Y'' is *der zur Zerlegung gehörende schwache Zerlegungsraum* (cf. Alexandroff and Hopf, p. 66). The *points* of Y'' are the *elements* of Y'.

(27) That is, the mapping, $y''=\alpha''(\alpha^{-1}(y))$, of Y on Y'' is a homeomorphism (cf. Alexandroff and Hopf, p. 61).

Proof. By Theorems 14.7 and 12.1 X and Y determine Hausdorff spaces. Let two of the conditions of Theorem 15.4 hold. The third follows; by Theorems VIII and II, pages 98 and 95, Alexandroff and Hopf, the mappings of (B) are equivalent and both are closed, and Y' is upper semi-continuous in X.

By Theorem 15.3 our hypothesis that X is upper semi-continuous relative to S is equivalent to condition (2) of Theorem 15.4. If $y=\alpha(x)$ is a closed mapping, condition (3) of that theorem holds. By the preceding paragraph the conditions in (A) and in (B) are necessary. Let $y''=\alpha''(x)$ be a closed mapping from X to Y''. Let $y \in Y$, U be an S-neighborhood of y, and $U(X)$ and $U(Y)$ denote all elements of X and of Y, respectively of which U is an S-neighborhood; $U(X)$ is open relative to X, and $U(X) \supset \alpha^{-1}(y) = y''$. If V is the set of all elements of Y'' that are subsets of $U(X)$, then V is an open set in Y'', and $y'' \in V$ (Definition (III)). Let H be the sum of the elements of $Y''-V$. Then y'' and H are closed point sets in X which do not intersect (cf. Theorem I, p. 53, Alexandroff and Hopf). Since X is perfectly compact, so are y'' and H. By Theorems 14.10, 14.7, and 11.6 y'' and H can be separated by S-neighborhoods D and R. Since $y''=\alpha^{-1}(y)$, $y<y''<y$; by Theorem 11.6 D is an S-neighborhood of y. Let D be an S-neighborhood of a portion β of an element of Y, y_1. Suppose that $y_1 \in Y-U(Y)$; by Theorem 11.6 there exists $x_1 \in \alpha^{-1}(y_1)$ such that U is not an S-neighborhood of x_1. Then $\alpha^{-1}(y_1) = \alpha''(x_1) \in Y''-V$, and $H \supset \alpha^{-1}(y_1)$. By Theorem 11.6 R is an S-neighborhood of y_1 and of β. Since D is an S-neighborhood of β, and $R \cdot D$ is vacuous, we are involved in a contradiction. Thus, Y is upper semi-continuous relative to S. By Theorems 15.3 and 15.4 the condition in (B) is sufficient. If the collection Y' is upper semi-continuous in X, it follows from Theorems VIII and II on pages 98 and 95, Alexandroff and Hopf, that the mapping $y''=\alpha''(x)$ is closed from X to Y''. By (B) the mapping $y=\alpha(x)$ is closed. By the first paragraph the condition in (A) is sufficient, and (C) is true.

THEOREM 15.6. *The conclusions of Theorems* 15.3, 15.4, *and* 15.5 *remain true if the aggregate consisting of P and of the elements of X and Y include regular real or regular frontier points.*

This follows from Theorems 12.1, 13.7, and 14.2.

16. **Applications of the preceding sections.** Applications of the preceding section require the decomposition of an amalgamation point P into regular amalgamation points. In problems dealing with the embedding of S in a perfectly compact Hausdorff space we have the case where P is the maximal amalgamation point and the elements of the decomposition include the points of S (cf. Theorem 16.1).

We supplement the results of §§13 and 14 for the regular, the normal, and the semi-completely normal spaces by giving characterizations of the completely regular, the locally perfectly compact, and the perfectly compact

Hausdorff spaces. Some of these give our interpretations of results by other authors, and demonstrate the generality of our methods (cf. Theorems 16.1, 16.2, and 16.4). In order to utilize this generality fully, we devote Chapter III to a study of lattices of regular points, and to lattices of systems of decompositions of an amalgamation point into collections of regular amalgamation points. These give systematic methods for classifying our results. An important part of Chapter III is the demonstration of the existence of atomic elements of systems of points which are encountered in applications of our mapping theory. For the special case of the semi-completely normal space the required atomic elements are the atomic ideal points (cf. Theorem 16.5). Similar conclusions hold for the decompositions of *boundary* points of *normal* spaces (cf. Theorem 16.1, (3)).

THEOREM 16.1. (1) *In order that S be completely regular*[28], *it is necessary and sufficient that there exist a collection of regular boundary points M such that* (a) *each boundary point intersect M, and* (b) *the collection $S+M$ be upper semi-continuous relative to S.* (2) *This condition is satisfied if and only if $S+M$ is a perfectly compact Hausdorff space.* (3) *If S is normal and M is the set of its atomic boundary points, this condition is satisfied.* (4) *A perfectly compact, immediate Hausdorff extension of S is a decomposition of the maximal ideal point.*

The result in (3) is similar to that in Lemma 12, p. 119, Wallman, loc. cit. If Y is any perfectly compact Hausdorff space in which S is embedded and X is the space $S+M$ of (3), then the conclusion of Theorem 15.4 concerning the mapping $y=\alpha(x)$ holds; cf. Stone, loc. cit., p. 476, Theorem 88. See also Theorem 20.2. Čech, also, loc. cit., has considered this space.

Proof. Let Y be the maximal amalgamation point and $X=S+M$. Let P be a real point, Q be a boundary point, and β be an atomic portion of Q. By (1a) β intersects $\beta_M \in M$; since β is atomic, $\beta < \beta_M$. Since β_M is regular, P and β_M do not intersect, and they can be separated by S-neighborhoods U and V (cf. D 13.1 and Theorem 11.5). Since $\beta < \beta_M$, V is an S-neighborhood of β and P and β do not intersect (Theorems 11.6 and 11.5). By Theorem 14.2 Q is an amalgamation point. By Theorem 14.6 P and Q do not intersect. By Theorem 13.1 P is regular. By (1a) each composition point intersects either S or M. By D 9.2 $Y < M+S$; since $M+S < Y$, $X=S+M$ is a decomposition of Y (cf. Theorem 15.2). By Theorems 14.5, 15.6, and 15.3 the condition in (1) is sufficient and that of (2) is necessary.

Conversely, let the perfectly compact Hausdorff space T be an immediate extension of S; then all points of T are regular. Let M be the boundary points of S which are equivalent to points of $T-S$. By Theorems 13.7 and 12.1,

(28) S is completely regular if and only if it can be embedded in a perfectly compact Hausdorff space; cf. Tychonoff, loc. cit.

$S+M$ is perfectly compact. By Theorems 14.10 and 14.2 there exists an amalgamation Z of the elements of $S+M$. If $Z\neq Y$, there exists an atomic boundary point β which intersects Y but does not intersect Z (cf. Theorem 9.2). By Theorems 12.1 and 14.7 the space $S+M+\beta$ is a Hausdorff space; clearly, β is a limit point of S. Since $S+M$ is perfectly compact, we are involved in a contradiction; cf. Alexandroff and Hopf, p. 91, Theorem XI. Thus, $Y<Z$, $Z<Y$, and $Z=Y$. By Theorems 15.6 and 15.3 the condition in (1) is necessary, and that of (2) is sufficient.

For Case (3) the collection $X=S+M$ obviously is upper semi-continuous. By Theorems 5.5 and 14.12 M satisfies condition (1a).

THEOREM 16.2. (1) *If S is regular but is not perfectly compact, then in order that the set of all boundary points be amalgamable it is necessary and sufficient that S be locally perfectly compact*; (2) *if this condition is satisfied and Q is the amalgamation of all the boundary points, then $S+Q$ is perfectly compact.*

This theorem resembles closely one due to Alexandroff; cf. Alexandroff and Hopf, p. 93, Theorem XIV.

Proof. If the condition holds, it follows from Alexandroff's theorem that there exists a point Q, such that $S+Q$ is a perfectly compact Hausdorff space in which S is embedded; because of Theorems 13.7 and 12.1 we may suppose that Q is a boundary point; by Theorem 14.2 Q is an amalgamation point. It follows from Theorem 16.1 that each boundary point is a portion of Q. Thus, the condition is sufficient.

Conversely, let Q be the amalgamation of all the boundary points. By Theorems 13.1 and 14.6 Q and the real point P can be separated by S-neighborhoods U and W, respectively. If $\overline{W}$ were not perfectly compact, by Theorem 10.3 there would exist a boundary point β which is an end of $\overline{W}$. Then β is a portion of Q, and U is an S-neighborhood of β. By Theorem 12.3 β is a limit point of $\overline{W}$, and $W\cdot U$ is nonvacuous. Thus, we are involved in a contradiction, and the condition is necessary.

THEOREM 16.3. *In order that a Hausdorff space be perfectly compact, it is necessary and sufficient that it have no boundary points.*

This follows from Theorem 10.3.

THEOREM 16.4. *Let X be a perfectly compact Hausdorff space which is an immediate extension of S, P be the maximal S-portion, M be a decomposition of X into closed point sets, and Y be the aggregate of all ideal points, $y(m)$, where $y(m)$ is the amalgamation of the elements of m, and $m\in M$. Theorems* 15.4 *and* 15.5 *are applicable.*

Thus, Y is upper semi-continuous relative to S if and only if M is upper semi-continuous in X; etc. The theorem points out that the results of Alexandroff and Hopf are special cases of our's.

THEOREM 16.5. *If S is semi-completely normal, the results of Theorems* 15.4 *and* 15.5 *hold true for the case that X is the decomposition of P into atomic ideal points.*

An interesting case is the one where P is the maximal ideal point, and Y is a perfectly compact Hausdorff space in which S is embedded. This justifies our regarding X as a *universal* inverse mapping space.

17. **Extensions of upper semi-continuous collections of point sets in normal spaces.** Stone has commented on the "remarkable properties" of a space which for the case that S is normal is a homeomorph of S plus its atomic boundary points (loc. cit., p. 476, lines 8 and 9). The results of this section give additional grounds for this comment. Our results in this section are distinguished by the fact that they are *characteristic* of normal spaces. Also, cf. Čech, loc. cit.

D 17.1. If S is a subspace of T, M and N are collections of mutually exclusive closed point sets of S and of T, respectively, and M is the collection of all sets $S\cdot n$, where $n\in N$, we say that N is *an extension of M (from S into T).*

E 17.1. Let T be a circle plus its interior, and S be the interior. Let N be the set of all chords parallel to a given diameter; define M as in D 17.1. N contains an extension of M, N is upper semi-continuous in T, but M is *not* upper semi-continuous in S. Thus, T cannot take the place of the space $\lambda(S)$ of Theorem 17.1.

Let E and F be two chords in T such that E and F have in common exactly one point, P, which belongs to $T-S$. Let K be the collection whose elements are $E\cdot S$, $F\cdot S$, and the points of $S-S\cdot(E+F)$. Then K is upper semi-continuous in S, but *cannot* be extended to T. It follows that T does not serve as a T_K (cf. Theorem 17.2).

THEOREM 17.1. *Let S be normal, $\lambda(S)$ be the space of S and its atomic boundary points, M be a collection of mutually exclusive closed point sets of S, N be the aggregate of the closures in $\lambda(S)$ of the elements of S, and K be the aggregate of the amalgamations of the elements of N.* (1) *The following conditions are equivalent*: (a) *M is upper semi-continuous in S*; (b) *N is upper semi-continuous in $\lambda(S)$*; (c) *K is upper semi-continuous relative to S.* (2) *If these conditions are satisfied, the spaces of these decompositions are homeomorphic*[29].

Thus, the theory of upper semi-continuous collections of point sets in a normal space may be regarded as a part of the theory of our ideal points (this holds true even if the elements of M are not perfectly compact).

Proof. For $m\in M$, let $n(m)\in N$ such that $m=S\cdot n(m)$; let $P(m)$ be the amalgamation of the elements of $n(m)$; by Theorems 16.1 and 14.10 $P(m)$ exists. If U is an open set in S, and $U\supset m\in M$, let $\lambda(U)$ denote all points of $\lambda(S)$ of which U is an S-neighborhood. Suppose that $\beta\in n(m)-m$ and not

(29) Cf. Definition (III), §15, and Theorem 12.1.

$\beta \in \lambda(U)$. Then by Theorems 11.3 and 12.3 β is an end of $S-U$ and of m. By Theorem 11.3 neither U nor $S-m$ is an S-neighborhood of β. Since $S=U+(S-m)$ and S is an S-neighborhood of β, by Theorem 11.7 we are involved in a contradiction. Thus, U is an S-neighborhood of $n(m)$ and of $P(m)$; cf. Theorem 11.6. Conversely, if W is an open set in $\lambda(S)$, and $W \supset n(m)$, then $n(m)$ and $\lambda(S)-W$ are closed in $\lambda(S)$. Thus, $n(m)$ and $\lambda(S)-W$ are perfectly compact in themselves and are amalgamable (cf. Theorems 16.1 and 14.10); by Theorems 14.7 and 11.6 they can be separated by S-neighborhoods, D and R. Then $W \supset \lambda(D) \supset n(m)$, $D \supset m$, and D is an S-neighborhood of $P(m)$. By Theorem 14.10 each $P(m)$ is regular. The conclusion follows from these relations, the definitions of upper semi-continuity, and Theorem IV, p. 53, Alexandroff and Hopf.

THEOREM 17.2. (1) *In order that it be possible to extend each M, which is an upper semi-continuous decomposition of S into closed point sets, to a subset of a similar decomposition of some immediate, perfectly compact, Hausdorff extension of S, say T_M, it is necessary and sufficient that S be normal.* (2) (a) *If S is normal, there exists a T_M, Y, which is independent of M*; (b) *such a Y is homeomorphic to $X=\lambda(S)$ by means of the mapping $y=\alpha(x)$ of Theorems* 15.6 *and* 15.4.

In particular, the conclusion of (2) holds for $Y=\lambda(S)$; cf. Theorem 17.1.

Proof. Let E and F be mutually exclusive closed sets in S. Let M be the aggregate whose elements are E, F, and the points of $S-(E+F)$. Then M is an upper semi-continuous decomposition of S. If a T_M exists, there exists an upper semi-continuous extension of M into T_M. Then the product of the closures of E and F in T_M is vacuous. Since T_M is perfectly compact, it is normal (cf. Alexandroff and Hopf, p. 89). There exist in T_M mutually exclusive open sets, U and V, which contain $\bar{E}_T$ and $\bar{F}_T$, respectively. Since $U \cdot S \supset E$ and $V \cdot S \supset F$, the condition in (1) is necessary.

Conversely, let S be normal and T be the space of the decomposition M (cf. Definition (III), §15). If $t \in T$ and $s \in S$, let the relation $t=f(s)$ mean that $s \in t$ (recall that the points of T are the elements of M). Then the mapping of S on T, $t=f(s)$, is continuous and T is normal (cf. Alexandroff and Hopf, pp. 67, 53, and 70). By Theorem 16.1 there exists an immediate, perfectly compact, Hausdorff extension of T, say R. There exists a continuous mapping from $\lambda(S)$ to R, say $r=F(s)$, such that if $P \in S$, then $f(P)=F(P)$ (cf. Theorem 16.1 and Stone, loc. cit., Theorem 88, p. 476).

Let N be the aggregate $[F^{-1}(r)]$, where r ranges over R. By Theorem VIII, p. 98, Alexandroff and Hopf, N is an upper semi-continuous decomposition of $Y=\lambda(S)$ into closed point sets. Thus, we have established (2a) and the sufficiency of the condition in (1).

Consider any Y which is independent of M, and let X be the space consisting of S and those boundary points which are equivalent to points of Y; the mapping of the conclusion is a homeomorphism of X and Y (cf. Theorem

13.7). Suppose that the element δ of $X-S$ is not atomic; then there exist two atomic boundary points which are portions of it, say α and β. By Theorem 14.12 α and β are regular. By Theorem 14.7 there exist S-neighborhoods of α and of β whose closures in S, E and F are mutually exclusive. If M is the collection whose elements are E, F, and the points of $S-(E+F)$, it is an upper semi-continuous decomposition of S. Clearly, α is a limit point of E and β is a limit point of F; by Theorem 11.6 each S-neighborhood of δ contains points of E and of F, and δ is a limit point of E and of F. Hence, there exist no mutually exclusive closed sets in Y which contain E and F, respectively. Since Y contains an extension of M, we are involved in a contradiction. Thus, δ does not exist, and $X=\lambda(S)$.

CHAPTER III. THE LATTICE-MAPPING THEORY OF SYSTEMS OF REGULAR POINTS

Let P be a regular amalgamation point, and let X, Y, and $y=\alpha(x)$ be such that the conditions of Theorem 15.4 are satisfied. Recall that for $y\in Y$, $X\supset\alpha^{-1}(y)$ and y is an amalgamation of the elements of $\alpha^{-1}(y)$. Since the mapping $y=\alpha(x)$ deals both with order and with continuity, we may think of Y as a mapping space or an amalgamation space for X, and of X as an inverse mapping space or a decomposition space for Y. If S is semi-completely normal, we have shown in Theorem 16.5 that if X^* is the decomposition of P into atomic ideal points, then X^* serves as a common inverse mapping space for all the Y's which are upper semi-continuous, or perfectly compact, decompositions of P into regular amalgamation points. That is, the elements of X^* may be said to be *atomic* from the point of view of our *mapping theory*. By Theorem 14.12 for spaces other than the semi-completely normal the atomic mapping points will not, in general, be the atomic ideal points. For these more general spaces there is thus the question of the existence of such atomic elements, and that of the existence of decomposition spaces or of inverse mapping spaces.

In §19 we give conditions for which these questions have answers in the affirmative. Our methods involve an order relation $X<Y$, where X and Y are decompositions of P which satisfy the conditions of Theorem 15.4. The ordered set so obtained is a complete lattice and its zero is the required set of atomic mapping points. The sublattices of this lattice, and their zeros and units, give an extensive body of information which is not in Chapter II for the case even of the semi-completely normal space. Important sublattices are considered in §20.

Section 18 is concerned with preliminary methods dealing with multiplicative systems and lattices of ideal points, and with the generation of such systems from collections of points. The zeros of the sublattices of §19 are, in general, collections of atomic elements of systems of ideal points.

In Example 14.3 we established the existence of a space for which the only

regular amalgamation point is the maximal ideal point P; for it $P=X=Y$, and the application of Theorem 15.4 is extremely limited.

18. **The generation of multiplicative systems and lattices of amalgamation points.** We develop methods for generating such systems from any collection of amalgamation points. Most of these do not require a hypothesis of regularity. In Theorem 18.5 we give conditions for regularity. In Theorem 18.4 we give conditions which make the generated set a lattice.

Note that if N is a subsystem of M, an atomic element[30] of N need not be an atomic element of M (for instance, N might consist of a single non-atomic element of M). The term *atomic regular* (*amalgamation*) *point* means an atomic element of the system of all regular amalgamation points; it should not be confused with the term regular atomic point (of S). Theorems 14.2 and 14.12 show that the terms are synonymous only for the semi-completely normal spaces.

D 18.1. The set of points M is said to be *almost-multiplicative* provided that if $M\supset K$ and K has a lower bound in the system of all ideal points, then the greatest lower bound of K belongs to M. The following are examples of almost-multiplicative systems: all atomic points; all composition points; all amalgamation points; all regular composition points; if S is completely regular, all regular amalgamation points (cf. Theorems 5.4, 14.2, 14.11, and 20.1).

D 18.2. If M is almost-multiplicative but is not completely multiplicative[31], it becomes a complete multiplicative system $M+\overline{O}$ by the addition of a zero element $\overline{O}$; if M is completely multiplicative, let $M+\overline{O}$ denote M, and let $\overline{O}$ denote its zero.

These systems find applications, among others, in establishing the existence of atomic elements. If M is the system of all regular amalgamation points of the space of Example E 14.3, it has one element, the maximal ideal point of the space. This is the atomic element of the system. This example shows how far the atomic regular points may differ from the atomic points.

THEOREM 18.1. *Let M be an almost-multiplicative system of amalgamation points.* (1) *If $P\in M$, there exists an atomic element of M which is a portion of P*; (2) *no two atomic elements of M intersect*; (3) *M is completely multiplicative if and only if it has exactly one atomic element* (*its zero*).

Proof. By Hausdorff, (I), pp. 140–141, there exists a system of elements of M, say K, which contains P, is monotonic, and is not a proper subset of any monotonic subcollection of M. For $k\in K$ the set $\alpha(k)$ is perfectly compact in itself (cf. Theorems 9.1 and 14.4). By Theorem 9.2 the $\alpha(k)$'s are the elements of a monotonic collection of point sets. There exists an atomic point δ which is common to all the $\alpha(k)$'s (cf. Kuratowski and Sierpiński, and Moore,

(30) Cf. D 3.3.

(31) Cf. D 3.7, and MacNeille, loc. cit., pp. 429, 442, and 443.

(II)). It follows that δ is a portion of each element of K. By D 18.1 there exists an element of M, β, which is the greatest lower bound of the elements of K. Let λ be an element of M such that $\lambda<\beta$, and let $k\in K$. Then $\lambda<\beta<k$. It follows that $K+\lambda$ is monotonic; because of the definition of K, $\lambda\in K$; then $\beta<\lambda$; since $\lambda<\beta$, $\lambda=\beta$, and β is an atomic element of M. Since $P\in K$, $\beta<P$.

Let α and β be intersecting atomic elements of M. By Theorem 8.1 and D 18.1 there exists an element of M, δ, which is a lower bound of α and β. Since $\delta<\beta$ and β is atomic, $\beta<\delta$, and $\beta=\delta$; similarly, $\delta=\alpha$. The conclusion of (2) and the necessity of the condition in (3) follow. If M has exactly one atomic element, it follows by (1) that this element is a lower bound of each collection of elements of M; by D 18.1 the condition in (3) is sufficient.

D 18.3. If K and $M(K)$ are collections of amalgamation points, $M(K)$ is an almost-multiplicative system, and $M(K)$ consists of those amalgamation points which are greatest lower bounds of subcollections of K, we say that $M(K)$ *is generated by* K; if $A(K)$ is the aggregate of all amalgamation points P such that P is the amalgamation of the elements of some subcollection of $M(K)$, we say that $A(K)$ *is the additive system generated by* K. If $x\in K$, x is the lower bound of the pair x and x; thus $M(K)\supset K$; similarly $A(K)\supset M(K)$. If K is either the set of all composition points, or of all atomic ideal points, then $K=M(K)$; if S is semi-completely normal, $A(K)$ is the set of all amalgamation points.

Theorem 18.2. *Each collection of amalgamation points K generates an additive system $A(K)$ and an almost-multiplicative system $M(K)$. $A(K)$ is an almost-multiplicative system, and $A(K)$ and $M(K)$ have the same atomic elements.*

Proof. Let $M(K)$ and $A(K)$ be defined as in D 18.3. Let H be a subcollection of $M(K)$ that has a lower bound. Let $H=H_1+H_2$ such that $K\supset H_1$ and $M(K)-K\supset H_2$. Let F be the product of all sets $\alpha(h)$, where $h\in H$ (cf. D 8.5). For $h\in H_2$ let $H(h)$ be a subcollection of K such that h is the greatest lower bound of the elements of $H(h)$; cf. D 18.3. Let H_3 be the sum of all the $H(h)$'s, for $h\in H_2$. Let $G=H_1+H_3$, and let E be the product of all the $\alpha(g)$'s, where $g\in G$. Clearly, $E\supset F$. Suppose that $\beta\in E-F$; there exists $h\in H$ such that β is not an element of $\alpha(h)$. Since $G\supset H_1$, $h\in H_2$; by Theorem 9.2 there exists $h_3\in H(h)$ such that $\alpha(h_3)$ does not contain β. Since $h_3\in G$, we are involved in a contradiction, and β does not exist. Thus, $E=F$; by Theorem 9.2 $P(F)$ is the greatest lower bound of the elements of H and of G. Since $K\supset G$, it follows from the definition of $M(K)$ that $P(F)\in M(K)$ (by Theorem 14.11 $P(F)$ is an amalgamation point). Thus, $M(K)$ is an almost-multiplicative system of amalgamation points, and is generated by K.

Next let H be a subcollection of $A(K)$, and let F be the product of all the sets $\alpha(h)$, for $h\in H$. If $h\in H$, there exists a subcollection of $M(K)$, M_h, such that h is the amalgamation of the elements of M_h (cf. D 18.3). For $\beta\in F$ and $h\in H$, let $x(\beta, h)$ be an element of M_h that intersects β. For a fixed β, let $x(\beta)$

be the greatest lower bound of the $x(\beta, h)$'s, where h ranges over H; by the first paragraph, $x(\beta)\in M(K)$. By Theorem 9.2 $F\supset\alpha(x(\beta))$, $x(\beta)<P(F)$, and $P(F)$ is the greatest lower bound of the elements of H. Let X be the aggregate of all the $x(\beta)$'s, for $\beta\in F$. By Theorems 8.1, 9.2, and D 9.2, $X<P(F)$, and $P(F)<X$. By Theorem 14.11 and D 14.2 $P(F)$ is an amalgamation of the elements of X. Since $M(K)\supset X$, it follows that $P(F)\in A(K)$, and that $A(K)$ is an almost-multiplicative system.

Since $A(K)\supset M(K)$, and for $a\in A(K)$ there exists $m\in M(K)$ such that $m<a$, it follows that the two systems have the same atomic elements.

THEOREM 18.3. *Let K be a system of amalgamation points, and $M(K)$ be the almost-multiplicative system generated by it. Then K and $M(K)$ generate the same additive system; and, if K is almost-multiplicative, $M(K)=K$.*

THEOREM 18.4. *If $A(K)$ is the additive system generated by the collection of amalgamation points K, then $A(K)+\overline{0}$ is a complete lattice if and only if it is possible to amalgamate the elements of K. If the condition is satisfied, the unit of the lattice is the amalgamation of the elements of K.*

Proof. Suppose that $A(K)+\overline{0}$ is a complete lattice; then it has a unit, I, which is an element of $A(K)$. Since $A(K)\supset K$, for $k\in K$, $k<I$; by D 9.2, $K<I$. Since $I\in A(K)$, it follows from D 18.3 and D 14.2 that there exists a subset of $M(K)$, say M, such that $I<M$. Let X be an ideal point that intersects I; by D 9.2 X intersects an element of M, say m. By the definition of $M(K)$ there exists $k\in K$ such that $m<k$. Then by D 9.2 X intersects k, and $I<K$; since $K<I$, it follows from D 14.2 that I is an amalgamation of the elements of K.

Conversely, let I be an amalgamation of the elements of K. Since $M(K)\supset K$, $I\in A(K)$. Let $\beta\in A(K)$; then there exists a subcollection of $M(K)$, M, such that $\beta<M<\beta$. For $m\in M$ there exists $k\in K$ such that $m<k$; by D 9.2 $M<K$; then $\beta<M<K<I$. Thus I is the unit of $A(K)$. By Theorem 18.2 and D 18.2 $A(K)+\overline{0}$ is a complete multiplicative system; since it has a unit, it is a complete lattice (cf. MacNeille, pp. 430–431).

THEOREM 18.5. *If K is a collection of regular amalgamation points, and each element of K is a portion of some collection of regular composition points, then the additive and the multiplicative systems generated by K consist of regular points.*

Proof. If $k\in K$ and $\beta\in\alpha(k)$ it follows from the hypothesis and Theorem 9.2 that β is a portion of some regular composition point X. By Theorems 9.2, 14.2, 14.11, and 14.8 the greatest lower bound of k and X is a regular composition point $k(\beta)$, and $\beta<k(\beta)$. It can be shown with the help of Theorem 9.1 that $k<\alpha(k)<L<k$, where L is the aggregate of all $k(\beta)$'s. Thus, k is an amalgamation of a collection of regular composition points. The conclusion follows from D 18.3 and Theorems 14.11 and 14.10.

19. Systems of upper semi-continuous, perfectly compact decompositions of regular amalgamation points. Let us consider the aggregate $D(R)$ of all decompositions of the perfectly compact Hausdorff space R into upper semi-continuous collections of closed point sets. If X and Y are elements of $D(R)$ and each element of X is a subset of some element of Y, let us say that $X<Y$. This relation orders $D(R)$. The following analogue to Theorem 19.1 holds: If Y_1 and Y_2 are elements of $D(R)$ and X is the set of all products $y_1 \cdot y_2$, where $y_1 \in Y_1$ and $y_2 \in Y_2$, then X is the greatest lower bound of Y_1 and Y_2 in $D(R)$. Then X is upper semi-continuous, and it corresponds to the $\omega_M(P)$ of Theorem 19.1, if M is the pair Y_1 and Y_2. Such analogies may be extended to the case of arbitrary subcollections M of $D(R)$. We shall now extend these ideas to the case of the amalgamation points.

D 19.1. If P is a regular amalgamation point, let $\delta(P)$ be the set of all δ's, where δ is an upper semi-continuous decomposition[32] of P into regular amalgamation points. If X and Y are elements of $\delta(P)$, let the relation $X<Y$ mean that each element of Y is decomposible into a set of points which is a subcollection of X; let $\delta(P)$ be ordered by this relation. Let $L(P)$ denote the sum of the elements of $\delta(P)$, and $\omega(P)$ be the set of atomic elements of $L(P)$.

Clearly this relation partially orders $\delta(P)$. The relation $X<Y$ is merely the requirement that X and Y satisfy the conditions of Theorem 15.4. Thus, the study of $\delta(P)$ and its sublattices systematizes our information about the mappings and inverse mappings we considered in §§15 and 16. We shall show that $\delta(P)$ is a complete lattice and that $\omega(P)$ is its zero. Thus, $\omega(P)$ may be mapped on any element of $\delta(P)$ by the methods of Theorem 15.4; the elements of $\omega(P)$ may be regarded as the *atomic points* from the point of view of these mappings. The zero of a sublattice of $\delta(P)$ has an analogous relation to the elements of the sublattice; and the elements of this zero may be regarded as the atomic points of the mapping theory which involves the elements of this sublattice.

EXAMPLE. E 19.1. Let P be an amalgamation point in a semi-completely normal space S or a boundary point of a normal space. Then $L(P)$ consists of all amalgamation points which are portions of P, and $\omega(P)$ is the decomposition of P into atomic points (cf. Theorems 16.1, 16.5, and 14.13).

D 19.2. If P is the maximal amalgamation point, let $\delta(S)=\delta(P)$ and $\omega(S)=\omega(P)$.

The preceding example suggests questions which the author has not solved for the case of the completely regular spaces. (I) If A and B are regular amalgamation points and $A<B$, does $\omega(B)\supset\omega(A)$? (II) Does $\omega(S)$ consist of the atomic regular points? An affirmative for (II) implies one for (I). A negative for (I) would imply the possibility of the existence of "incommensurable

(32) By Theorem 15.3 the condition of the upper semi-continuity of δ is equivalent to that of its perfect compactness (cf. D 9.3).

points," A and B; or that of a point P and a point β such that $\beta \in \omega(P)$ and $\omega(\beta)$ consists of proper portions of β. Theorem 14.12 shows that if S is not semi-completely normal, some of its atomic points are too fine to be regular; the preceding discussion suggests that for more general spaces S some of the atomic regular points may be too fine to belong to $\omega(S)$. In any case the existence of $\omega(P)$ is of importance, since it is the finest decomposition of P for which the methods of §15 are applicable.

THEOREM 19.1. *Let P be a regular amalgamation point which is a portion of a collection of regular composition points; let M be a subcollection of $\delta(P)$; let $L_M(P)$ be the additive system which is generated by the sum of the elements of M, and $\omega_M(P)$ be the set of atomic elements of $L_M(P)$: (1) $\omega_M(P) \in \delta(P)$; (2) if $\beta \in \omega_M(P)$ and $N \in M$, β intersects exactly one element of N, say N_β; and β is the greatest lower bound of all the N_β's; (3) $L_M(P)$ is the sum of all the Y's such that $Y \in \delta(P)$ and $\omega_M(P) < Y$ in $\delta(P)$; the elements of M are such Y's; (4) $L_M(P)+\overline{0}$ is a complete lattice of regular amalgamation points, and its unit is P.*

The lattice of (4) is obtained by the methods of the preceding section. The set of its atomic elements, $\omega_M(P)$, is the zero of a certain sublattice of $\delta(P)$; this lattice is suggested in condition (3) (cf. Theorem 19.3). This condition is used to show that the systems in the following two theorems are complete multiplicative systems.

Proof. Let β be an atomic portion of P; it intersects exactly one element, N_β, of each element N of M. By Theorem 9.2, $\beta < P_\beta$, the greatest lower bound of all the N_β's. By D 18.3 $P_\beta \in L_M(P)$; by Theorem 18.5 P_β is regular. By Theorems 18.1 and 18.2 there exists an atomic element of $L_M(P)$, say A, which is a portion of P_β; and there exists a subcollection of the sum of the elements of M, say A^*, such that A is the greatest lower bound of A^*. Since $A < P_\beta$, A is a lower bound of the elements of $[N_\beta]$. If $N \in M$, no two elements of N intersect, since N is a decomposition of P into amalgamation points. It follows that $[N_\beta] \supset A^*$; by Theorem 9.2 $P_\beta < A$; since $A < P_\beta$, $A = P_\beta$; thus we have established (2). Since $\beta < P_\beta \in \omega_M(P)$, it follows with the help of Theorems 9.1 and 9.2 that $P < \alpha(P) < \omega_M(P) < P$; by Theorem 18.1 no two elements of $\omega_M(P)$ intersect; thus $\omega_M(P)$ is a decomposition of P. It remains to show that $\omega_M(P)$ is upper semi-continuous.

First we shall establish this upper semi-continuity for the case that the number λ of elements of M is a positive integer. Let $n = \lambda$ be a positive integer such that the conclusion holds for each collection of n elements of $\delta(P)$; clearly, $n = 1$ is such an integer. Let M and H be two subcollections of $\delta(P)$ which have, respectively, $n+1$ and n elements, and are such that $M \supset H$. Let $J = \omega_H(P)$, K be the element of $M - H$, and L be the set of atomic elements of the additive system generated by $J + K$. It follows from part (2)

and Theorem 9.2 that $L=\omega_M(P)$. Let $\beta\in L$ and let D be an S-neighborhood of β. There exist elements of J and K, say j and k, such that β is the greatest lower bound of j and k. By an argument used in the proof of Theorem 14.11 there exist in S mutually exclusive open sets D_1 and D_2 such that $D+D_1$ and $D+D_2$ are S-neighborhoods, respectively, of j and of k. Since J is upper semi-continuous, there exists an S-neighborhood of J, R_1, such that if R_1 is an S-neighborhood of a portion of an element of j, then $D+D_1$ is an S-neighborhood of that element. Similarly define R_2 for k and $D+D_2$; let $R=R_1\cdot R_2$. Let R be an S-neighborhood of a portion of an element of L, β_1, which is the greatest lower bound of j_1 and k_1. Then $D+D_1$ and $D+D_2$ are S-neighborhoods of j_1 and of k_1. Since $D_1\cdot D_2$ is vacuous, it follows from Theorems 9.2 and 11.6 that D is an S-neighborhood of $\alpha(\beta_1)=\alpha(j_1)\cdot\alpha(k_1)$ and of β_1. Thus L is upper semi-continuous, and our conclusion holds for any finite case.

Let M be infinite. Again let $\beta\in\omega_M(P)$ and D be an S-neighborhood of β. By (2) β is the greatest lower bound of the elements of $[N_\beta]$; by Theorem 9.2 $\alpha(\beta)$ is the product of the sets $\alpha(N_\beta)$. By methods used in the proof of Theorem 14.11 there exist a finite collection of the N_β's, $H=(N_{1\beta}, N_{2\beta}, \cdots, N_{k\beta})$ such that D is an S-neighborhood of the product $X=\alpha(N_{1\beta})\cdot\alpha(N_{2\beta})\cdots\alpha(N_{k\beta})$. If β_k is the greatest lower bound of the elements of H, by Theorems 9.2 and 11.6 D is an S-neighborhood of β_k. Let $N_{i\beta}\in N_i\in M$, $G=(N_1, N_2, \cdots, N_k)$, and $L=\omega_G(P)$. By the preceding paragraph $L\in\delta(K)$ and $\beta_k\in L$. Since L is upper semi-continuous, there exists an S-neighborhood of β_k, R, such that if R is an S-neighborhood of a portion of an element of L then D is an S-neighborhood of that element. By Theorem 9.2 $\beta<\beta_k$; by Theorem 11.6 R is an S-neighborhood of β. By condition (2) each element of $\omega_M(P)$ is a portion of an element of L; it follows that $\omega_M(P)$ is upper semi-continuous. We have established (1).

Let Z be the almost-multiplicative system which is generated by the sum of the elements of M. By Theorem 18.2 $\omega_M(P)$ is the set of atomic elements of Z. Let $Y\in\delta(P)$ and $\omega_M(P)<Y$ in $\delta(P)$; by D 19.1 the latter means that each element of Y, y, is an amalgamation of a subcollection of $\omega_M(P)$. Since $Z\supset\omega_M(P)$, $L_M(P)\supset Y$ (cf. D 18.3). Thus, if Σ is the sum of all Y's that satisfy the condition of (3), then $L_M(P)\supset\Sigma$. Conversely, let $A\in L_M(P)$ and let H be the set of all elements of $\omega_M(P)$ that are portions of A; then A is an amalgamation of the elements of H. Let L be the set which consists of H and the elements of $\omega_M(P)-H$; let Y consist of A and the elements of $\omega_M(P)-H$. Since $L_M(P)$ is an almost-multiplicative system and its atomic elements are the elements of $\omega_M(P)$, no element of $\omega_M(P)-H$ intersects A (cf. Theorems 18.1 and 18.2). By Theorem 14.7 A and H can be separated from each element of $\omega_M(P)-H$ by S-neighborhoods, and none of these elements is a limit point of H. Thus, H is closed relative to $\omega_M(P)$. By Theorem 15.3 $\omega_M(P)$ is perfectly compact in itself. Since L is an upper semi-continuous decomposition of $\omega_M(P)$ into closed point sets, it follows from Theorem 15.5 that Y is upper semi-con-

tinuous relative to S. Then $Y \in \delta(P)$ and $\omega_M(P) < Y$ in $\delta(P)$. Thus, $\Sigma \supset L_M(P)$; since $L_M(P) \supset \Sigma$, $L_M(P) = \Sigma$.

The conclusion of (4) follows from Theorems 18.4 and 18.5.

THEOREM 19.2. *Let P be the point of Theorem 19.1. If $M = \delta(P)$, then $L(P) = L_M(P)$ and $\omega(P) = \omega_M(P)$. Also, $\delta(P)$ is a complete lattice with unit P and zero $\omega(P)$.*

Proof. If $M = \delta(P)$, by Theorem 19.1 and D 19.1 $L_M(P) \supset L(P)$, and conversely. Thus, $L_M(P) = L(P)$ and $\omega_M(P) = \omega(P)$. By Theorem 19.1, (3), if $Y \in \delta(P)$, then $\omega(P) < Y$ in $\delta(P)$. Thus, $\omega(P)$ is the zero of $\delta(P)$; clearly P is the unit of $\delta(P)$.

By Theorem 19.1 if $\delta(P) \supset M$ and $Y \in M$, then $\omega_M(P) < Y$ in $\delta(P)$. Let X be a lower bound of M in $\delta(P)$. If $x \in X$ and $N \in M$, there exists one element of N, N_x, such that x is a portion of N_x (cf. D 19.1). By (2) of Theorem 19.1 there exists an element of $\omega_M(P)$, say β_x, which is the greatest lower bound of the N_x's. Thus, x is a portion of β_x. From this relation and the fact that each of X and $\omega_M(P)$ is a decomposition of P, each element of $\omega_M(P)$ can be decomposed into a subcollection of X. Thus, $X < \omega_M(P)$ and $\omega_M(P)$ is the greatest lower bound of M in $\delta(P)$. Since $\delta(P)$ has a unit, it is a complete lattice (cf. MacNeille, pp. 430, 431, and Birkhoff, p. 17).

THEOREM 19.3. *If $\delta(P) \supset M$ and P satisfies the condition of Theorem 19.1, let $\omega_1(M, P) = \omega(P)$, $\omega_2(M, P) = \omega_M(P)$, $\omega_3(M, P)$ be the least upper bound of M in $\delta(P)$, and $\omega_4(M, P) = P$; if $1 \leqq i \leqq j \leqq 4$, let $\delta_{ij}(M, P)$ be the set of all elements of $\delta(P)$, X, such that $\omega_i(M, P) < X < \omega_j(M, P)$ in $\delta(P)$. Then $\delta_{ij}(M, P)$ is a complete sublattice of $\delta(P)$ and its zero and unit are, respectively, $\omega_i(M, P)$ and $\omega_j(M, P)$; $\omega_M(P)$ is the greatest lower bound of M in $\delta(P)$.*

Thus, any $X < \omega_2(M, P)$ may serve as a common inverse mapping space for the elements of M. However, $X = \omega_2(M, P)$ may be regarded as the "most economical" of these inverse spaces. For, if $X \in \delta_{12}(M, P)$ and Y_1 and Y_2 are elements of M, we can first decompose the elements of Y_1 into points of X and then reamalgamate these into points of Y_2 by the methods of Theorem 15.4. Since the elements of $X = \omega_2(M, P)$ are "coarser" than those of any other element of $\delta_{12}(M, P)$, the initial decomposition need not be extended so far if this X is used for an inverse space. Similarly, $X = \omega(P)$ is the "least economical" of these inverse spaces, but it serves for all sets in $\delta(P)$. Analogous interpretations hold for the common mapping spaces of the elements of M, the elements of $\delta_{34}(M, P)$. Theorem 19.3 may be modified by replacing $\delta(P)$ by any of its complete sublattices.

20. **Applications to completely regular spaces.** If we order all the perfectly compact immediate Hausdorff extensions of S by the ordering of D 19.1, we obtain a quasi-partially ordered system (cf. Theorems 13.7 and 16.1). By

identifying[33] equivalent elements of this system, we obtain a complete multiplicative system which is isomorphic to the subsystem of $\delta(S)$, $H(S)$, which we consider in D 20.1 and Theorems 20.2 to 20.4. The zero of $H(S)$, $\lambda(S)$, is equivalent to the space considered by Stone in his Theorem 88, p. 476. In §17 we considered this space for the case that S is normal (cf. also, Čech, loc. cit.)

The important theorems 14.11, 18.5, and 19.1 to 19.3 involve hypotheses in which certain points satisfy the condition of being the amalgamation of a collection of regular composition points[34]. This condition and the results of the preceding section enter into the characterizations of completely regular spaces which are given in Theorem 20.1. The fact that the atomic mapping points in such a space are necessarily regular composition points gives an analogy to the results of Theorem 14.13.

D 20.1. Let $H(S)$ be the subsystem of $\delta(S)$ which consists of all Y's such that each point of Y is either a boundary point or is equivalent[35] to a point of S. Let $\lambda(S)$ be the zero of $H(S)$ (cf. D 19.2).

THEOREM 20.1. *Any of the following is a necessary and sufficient condition that S be completely regular*: (1) *At least one element of $\delta(S)$ consists of composition points*; (2) *the elements of $\omega(S)$ are composition points*; (3) *if Q is a regular amalgamation point, the elements of $\omega(Q)$ are regular composition points.*

Any regular amalgamation point in a completely regular space is the amalgamation of some collection of regular composition points[36].

Proof. By Theorems 13.7 and 16.1 the condition in (1) is necessary. Let Y be an element of $\delta(S)$ which consists of regular composition points, Q be a regular amalgamation point, and K be the aggregate of all points Q_y, where Q_y is the greatest lower bound of Q and the element y of Y. By Theorem 14.11 if y intersects Q, then Q_y is a regular amalgamation point. Since Y is a decomposition of the maximal amalgamation point P, and $Q<P$, it may be shown that K is a decomposition of Q. We may regard Q as an upper semi-continuous collection of one element; since Y is upper semi-continuous, we may show by an argument like that used for the finite case in the proof of Theorem 19.1 that K is upper semi-continuous relative to S. By D 19.1 $K\in\delta(Q)$. By Theorem 19.1 if $\beta\in\omega(Q)$, there exist elements of K and of Y, respectively, k_β and y_β, such that $\beta<k_\beta<y_\beta$. Since β is a regular amalgamation point and y_β is a regular composition point, it follows from Theorem 14.8 that β is a composition point. Thus, the conditions in (2) and (3) are necessary. Since $\omega(Q)\in\delta(Q)$, by D 19.1, D 14.2, and D 9.3 Q is an amalgamation of $\omega(Q)$.

---

(33) Cf. Birkhoff, p. 7, Theorem 1.2.

(34) There exist examples which show that such an hypothesis is essential for the truth of these theorems.

(35) Cf. D 3.2 and Theorem 13.7.

(36) Cf. Theorem 14.13.

Clearly, each of (2) and (3) implies (1) (cf. Theorem 14.5). Let X be an element of $\delta(S)$ which consists of regular composition points, M be the set of those elements of X which are boundary points, and $Y=S+M$. By Theorem 5.2 no boundary point intersects a decomposition point. By D 19.1, D 19.2, and D 9.2 each boundary point intersects M. By the argument used at the beginning of the proof of Theorem 16.1 it follows that the points of S are regular. Since no two elements of Y intersect, it follows from Theorems 14.6, 14.7, and 12.1 that Y is a Hausdorff space. Let the mapping $y=\alpha(x)$ be defined as in Theorems 15.4 and 15.6. Let $\beta\in Y$, and D be an S-neighborhood of β. Since β is regular, there exists an S-neighborhood of β, R, such that $D\supset\overline{R}_S$. Let x be an element of X such that R is an S-neighborhood of x. By Theorem 12.4 $\overline{R}_S\supset a(x)$. By D 12.2 and the definition of the mapping, $y=\alpha(x)$, either $x=\alpha(x)$ or $a(x)=\alpha(x)$; in either case D is an S-neighborhood of $\alpha(x)$. Thus, our mapping is continuous from X to Y (cf. Alexandroff and Hopf, p. 53, Theorem IV). By Alexandroff and Hopf, Theorem VIII, p. 98, Y is perfectly compact. By Theorems 15.3 and 16.1 the condition in (1) is sufficient.

THEOREM 20.2. *Let S be completely regular*: (1) *$H(S)$ is a complete multiplicative subsystem of $\delta(S)$*; (2) *$\lambda(S)$ consists of the composition points which are equivalent to the points of S, and of the boundary points which are elements of $\omega(S)$*; (3) *if $X=\lambda(S)$ and $Y\in H(S)$, then X may be mapped on Y by the mapping $y=\alpha(x)$ of Theorem* 15.4.

Concerning (3) cf. Stone, loc. cit., Theorem 88.

Proof. Let H' denote the set of composition points which are equivalent to points of H (cf. Theorem 13.7 and D 3.2). Let M be a subcollection of $H(S)$. By Theorem 19.3 $\omega_M(S)$ is the greatest lower bound of M in $\delta(S)$. Since S' is a subset of each element of M, and the remaining points of each element of M are boundary points, by Condition (2) of Theorem 19.1 $\omega_M(S)\supset S'$, and $\omega_M(S)-S'$ consists of boundary points (cf. Theorem 14.8). Thus, $\omega_M(S)\in H(S)$; (1) follows by D 3.7. If $M=H(S)$, by Theorem 19.3 and D 20.1 $\lambda(S)=\omega_M(S)$. Let K be the set of all boundary points which belong to $\omega(S)$, and let $Y=S'+K$. By an argument similar to that used in the proof of the sufficiency of the condition of Theorem 20.1 it follows that $Y\in\delta(S)$; hence, $Y\in H(S)$. Thus, $\lambda(S)<Y$ in $H(S)$. Since $\omega(S)<\lambda(S)$ in $\delta(S)$, it follows that $\lambda(S)-S'=K$, and that $\lambda(S)=Y$.

D 20.2. Let the symbol H' denote the set of all composition points which are equivalent to elements of H (cf. D 3.2 and Theorem 13.7). Let B be the set of all boundary points and C be the set of all real points which are limit points of B. Let $Q(S)$ be the amalgamation of the elements of $B+C$. Let $I(S)=(S'-C')+Q(S)$.

Recall that $H(S)$ contains the topological images of all immediate, perfectly compact Hausdorff extensions of S. Theorem 20.2 shows that $\lambda(S)$ is a universal inverse mapping space for such extensions, and is one of them. The

two following theorems show that $I(S)$ is the least common mapping space of the elements of $H(S)$, and that it does not necessarily belong to $H(S)$. The following theorem gives the structure of $I(S)$. In particular, if S is not locally perfectly compact at any of its points, then $Q(S)$ is $I(S)$ and is the maximal amalgamation point. Then $Q(S)$ is the unit of $\delta(S)$.

THEOREM 20.3. *Let S be completely regular.* (1) *$I(S)$ is the least upper bound of $H(S)$ in $\delta(S)$*; (2) *$I(S)-Q(S)$ consists of those composition points which are equivalent to the points of S at which S is locally perfectly compact.*

Proof. Let M be the set of all boundary points, N be the set of real points which are limit points of M, and $K=S-N$. If $X\in K$, there exists an S-neighborhood of X, R, which is not an S-neighborhood of any element of $M+N$ (cf. Theorems 12.2 and 13.1). Let D be an S-neighborhood of X such that $R\supset\overline{D}$. If $\overline{D}$ were not perfectly compact, there would exist a boundary point β such that β is an end of $\overline{D}$ (Theorem 10.3). Then, since $R\supset\overline{D}$, it may be shown with the help of Theorem 10.1 that R is an S-neighborhood of β; since $\beta\in M$, we are involved in a contradiction. It follows that if the composition point δ is a limit point of $M+N$, then $X\neq a(\delta)$ (cf. D 12.2). Thus, either $a(\delta)\in N$, or $a(\delta)$ is vacuous, and δ is a boundary point; in either case δ intersects an element of $M+N$. By Theorems 14.1 and 13.7 $Q(S)$ is an amalgamation point. Clearly, no two elements of $I(S)$ intersect, each ideal point intersects an element of $I(S)$, and $I(S)$ is upper semi-continuous relative to S (cf. D 15.1 and D 19.2). Thus, $I(S)\in\delta(S)$.

Let $Y\in H(S)$ and Z be an upper bound of $H(S)$ in $\delta(S)$. Clearly, $Y<I(S)$ in $\delta(S)$. If A and B are boundary points which belong to elements of $H(S)$, they can be decomposed into subcollections of $\lambda(S)$, H_A and H_B; and they can be amalgamated into a point C such that $C<H_A+H_B<C$. Let $L=C+\lambda(S)-(H_A+H_B)$. Then C intersects no element of $L-C$; since $\lambda(S)$ contains $L-C$ and is upper semi-continuous, it follows that L is upper semi-continuous; also, L is a decomposition of the maximal amalgamation point. Thus, L is an element of $\delta(S)$ and of $H(S)$, and $L<Z$. Let $\delta\in Z$ such that C is a portion of δ; since A and B are arbitrary, it follows that δ is independent of A, B, and C; thus, by D 9.2 M is a portion of δ. Since the elements of N are limit points of M, they intersect δ (cf. Theorem 14.6). Since $S'\supset N'$, each element of S' belongs to each element of $H(S)$, and no two elements of Z intersect, each element of N is a portion of some element of Z; *that is, of* δ. By D 9.2 $Q(S)$ is a portion of δ. Since the elements of $I(S)-Q(S)$ belong to S', it follows that $I(S)<Z$ in $\delta(S)$. We have established the conclusion.

THEOREM 20.4. (1) *In order that $H(S)$ be a complete sublattice of $\delta(S)$, it is necessary and sufficient that S be locally perfectly compact.* (2) *If this condition is satisfied, $\lambda(S)$ and $I(S)$ are the zero and the unit of $H(S)$; and $Q(S)$ is the amalgamation of all the boundary points.*

Here $I(S)$ is the topological image of the space we considered in Theorem 16.2. Cf. Alexandroff and Hopf, p. 93.

Bibliography

Alexandroff, Paul, *Über stetige Abbildungen kompakter Räume*, Mathematische Annalen, vol. 96 (1926), pp. 555–571.

Alexandroff, P., and Hopf, H., *Topologie*, Berlin, Springer, 1935.

Alexandroff, P., and Urysohn, P., (I) *Zur Theorie der topologischen Räume*, Mathematische Annalen, vol. 92 (1924), pp. 258–266; (II) *Mémoire sur les espaces topologiques compacts*, Verhandelingen der Koninklijke Akademie van Wetenschappen, Amsterdam, vol. 14 (1929), pp. 1–96.

Birkhoff, Garrett, *Lattice Theory*, American Mathematical Society Colloquium Publications, vol. 25, New York, 1940.

Carathéodory, C., *Über die Begrenzung einfach zusammenhängender Gebiete*, Mathematische Annalen, vol. 73 (1913), pp. 323–370.

Cartan, H., (I) *Théorie des filtres*, Comptes Rendus de l'Académie des Sciences, vol. 205 (1937), pp. 595–598; (II) *Filtres et ultrafiltres*, ibid., pp. 777–779.

Čech, E., *On bicompact spaces*, Annals of Mathematics, (2), vol. 38 (1937), pp. 823–844.

Fréchet, M., (I) *Les Espaces Abstraits et Leur Théorie Considerée comme Introduction à l'Analyse Générale*, Paris, 1928; (II) *Sur les ensembles abstraits*, Annales de l'École Normale Supérieure, vol. 38 (1921), pp. 341–387.

Hausdorff, F., (I) *Grundzüge der Mengenlehre*, 1914; (II) *Mengenlehre*, 1927.

Jones, F. Burton, *Concerning normal and completely normal spaces*, Bulletin of the American Mathematical Society, vol. 43 (1937), pp. 673–675.

Kuratowski, C., *Topologie*, 1933.

Kuratowski, C., and Sierpiński, W., *Le théorème de Borel-Lebesgue dans la théorie des ensembles abstraits*, Fundamenta Mathematicae, vol. 2 (1921), pp. 172–179.

MacNeille, H. M., *Partially ordered sets*, these Transactions, vol. 42 (1937), pp. 416–460.

Menger, K., (I) *Dimensionstheorie*, 1928, and (II) *Kurventheorie*, 1932.

Moore, R. L., (I) *Foundations of Point Set Theory*, American Mathematical Society Colloquium Publications, vol. 13, 1932; (II) *On the most general class L Fréchet in which the Heine-Borel-Lebesgue theorem holds true*, Proceedings of the National Academy of Sciences, vol. 5 (1919), pp. 206–210; (III) *Concerning upper semi-continuous collections of continua*, these Transactions, vol. 27 (1925), pp. 416–428.

Sierpiński, W., *Introduction to General Topology*, 1934, translated by C. C. Krieger.

Stone, M. H., *Applications of the theory of Boolean rings to general topology*, these Transactions, vol. 41 (1937), pp. 375–481.

Tietze, H., *Beiträge zur allgemeinen Topologie*, Mathematische Annalen, vol. 88 (1923), pp. 290–312.

Tychonoff, A., *Über die topologische Erweiterung von Räumen*, Mathematische Annalen, vol. 102 (1929–1930), pp. 545–561.

Urysohn, P., *Über die Mächtigkeit der zusammenhängender Mengen*, Mathematische Annalen, vol. 94 (1925), pp. 263–295.

Wallman, H., *Lattices and topological spaces*, Annals of Mathematics, (2), vol. 39 (1938), pp. 112–126.

University of Texas,
Austin, Texas.

ON THE JACOBI SERIES

BY

J. H. CURTISS

1. Introduction. Let $F(z)$ denote a function regular in a neighborhood of each of the points of the complex z-plane determined by the λ numbers α_j, $j=1, 2, \cdots, \lambda$, which are not necessarily distinct. By means of interpolation to $F(z)$ in the points α_j, the coefficients a_n of the following series are uniquely determined [20, p. 53]:

$$\begin{aligned} &a_0 + a_1(z-\alpha_1) + a_2(z-\alpha_1)(z-\alpha_2) + \cdots + a_{\lambda-1}(z-\alpha_1)\cdots(z-\alpha_{\lambda-1}) \\ (1.1)\quad &+ a_\lambda\omega(z) + a_{\lambda+1}(z-\alpha_1)\omega(z) + \cdots \qquad + a_{2\lambda-1}(z-\alpha_1)\cdots(z-\alpha_{\lambda-1})\omega(z) \\ &+ a_{2\lambda}(\omega(z))^2 + \cdots, \end{aligned}$$

where $\omega(z)=\prod_{j=1}^{\lambda}(z-\alpha_j)$. We shall denote the partial sums of this series by $S_n(z; F)$ or $S_n(z)$, $n=0, 1, 2, \cdots$, and the Cesàro means of order r by $S_n^{(r)}(z; F)$ or $S_n^{(r)}(z)$, $n=0, 1, 2, \cdots$.

Jacobi [8](1) seems to have been the first to study developments of this type. He was interested in the problem of finding a formal expansion for $F(z)$ of the type $\sum_0^\infty q_\nu(z)(\omega(z))^\nu$, in which the functions $q_\nu(z)$ are polynomials of degree less than λ. The sum of the first n terms of this original series of Jacobi is the polynomial of degree at most $\lambda n-1$ which interpolates to the function $F(z)$ in the points α_j, each considered to be of multiplicity n. Thus the nth partial sum of Jacobi's original series is identical with the λnth partial sum of the series (1.1). A change in the order of the points α_j naturally changes (1.1), but does not change Jacobi's original series. We shall call (1.1) the Jacobi series for $F(z)$ with respect to the points α_j. The series is a generalization of the Taylor series, to which it reduces if the α_j all coincide.

The present study of the Jacobi series was undertaken at the suggestion of Professor J. L. Walsh. The purpose of the paper is to develop two general methods for studying the Jacobi series on the boundaries of its regions of convergence, and to obtain thereby certain typical results concerning the behavior of the series on these boundaries. The first method, in which the basic idea (§4) is due to Professor Walsh, consists in the study of certain expres-

Presented to the Society in two parts: §§1–5 and 9, December 27, 1934, under the title *The Jacobi interpolation series on the lemniscate of convergence* (jointly with Professor J. L. Walsh); §§10–11, October 30, 1937, under the title *Uniform convergence and summability of the Jacobi series on an unrestricted lemniscate*; received by the editors July 15, 1939, and, in revised form, June 13, 1940. The author is indebted to Professor Walsh not only for the contributions acknowledged in the text, but also for a number of suggestions concerning the exposition of the results contained in this paper.

(1) See also [20, pp. 54–64]; and [5, 6, 9, 10, 11, 12, 13, 19]. The Jacobi series has been used by Lebesgue to establish Weierstrass's theorem on approximation to continuous functions by polynomials; see [2, p. 60].

sions for the sums of the columns of the array (1.1). This method is developed in §§4–7, and is applied in §8 to the problem of determining the order of the coefficients of the series (1.1) under various conditions on the function $F(z)$. The second method consists in expressing the λnth partial sum of (1.1) (that is, the nth partial sum of Jacobi's original series) as a definite integral similar to Dirichlet's integral (§10). The technique of working with this integral is illustrated in §11 by the derivation of several convergence tests for the Jacobi series analogous to the de la Vallée Poussin test and related tests for Fourier series. Both of the two general methods depend upon a fundamental connection between the Jacobi series and the Fourier series of the boundary values of $F(z)$ (Theorem 7.7). The existence of such boundary values is discussed in some detail in §7, not only because it is an essential step in establishing the relation between the Jacobi and Fourier series, but also because it is a funcion-theoretic problem of some interest in itself.

Many further results concerning the convergence and summability of the Jacobi series are quite immediate consequences of the work in this paper. It is perhaps well to mention that one of the chief reasons for undertaking the present study was the hope that Jacobi series results might be indicative of those to be expected in the study of certain more general series of interpolation, and that the methods developed here might also be applicable in more general situations.

2. **The regions of convergence.** There exists a greatest positive number μ (finite or infinite) with the property that $F(z)$ is single-valued and regular on the set D: $|\omega(z)| <\mu$ [20, p. 58]. This set consists of the finite plane if $\mu=\infty$. If $\mu<\infty$, it consists of λ' $(1\leqq\lambda'\leqq\lambda)$ mutually exclusive Jordan regions D_k, $k=1, \cdots, \lambda'$, the boundaries of which are contours(2) which we shall denote respectively by the letters C_k, $k=1, \cdots, \lambda'$. The set D_k+C_k will be denoted by $\overline{D}_k$, and the set $|\omega(z)|\leqq\mu$ by $\overline{D}$. The set $\sum_1^{\lambda'} C_k$ is the lemniscate(3) Γ: $|\omega(z)|=\mu$, which, for reasons contained in Theorem 2.1 below, is called the lemniscate of convergence of the series (1.1). In the neighborhood of a point β on Γ at which $d\omega/dz=\omega'(z)$ has an $(m-1)$-fold zero, the locus Γ consists of m analytic arcs passing through the point β, with equally spaced tangents(4). The point β is called a multiple point of order m of Γ. By the index of a set E on the lemniscate Γ, we shall mean the number $m(E)$ such that $\omega'(z)$ has a zero of order $m(E)-1$ on E but has no zero of higher order on E. We shall always denote the index of Γ by $\bar{m}$.

The fundamental convergence theorem is as follows [20, pp. 57–60](5):

(2) By a contour, we mean a Jordan curve of the finite plane composed of a finite number of analytic Jordan arcs.

(3) For a complete discussion of lemniscates, the reader is referred to [20, pp. 54–56], and [21].

(4) This statement is easily proved by the implicit function theorem.

(5) See also [10] and [11].

THEOREM 2.1. *The series* (1.1) *converges absolutely for z on D, and uniformly to $F(z)$ for z on any closed limited subset of D. The series diverges for $|\omega(z)| > \mu$. Moreover, we have*

$$\limsup_{n\to\infty} |a_n|^{1/n} = (1/\mu)^{1/\lambda},$$

$$|F(z) - S_n(z)| \leqq M(\mu_1/\mu_2)^{n/\lambda},$$

$$|\omega(z)| \leqq \mu_1 < \mu,$$

where M is independent of n and z but not of μ_1 and μ_2, and where $\mu_2 > \mu_1$ is an arbitrary positive number less than μ.

The coefficients a_n are given by the formulas(6)

$$a_{\lambda\nu+K} = \frac{1}{2\pi i}\int_{\Gamma'} \frac{F(z)dz}{(z-\alpha_1)(z-\alpha_2)\cdots(z-\alpha_{K+1})[\omega(z)]^{\nu}}, \tag{2.1}$$

$$K = 0, 1, \cdots, \lambda - 1, \nu = 0, 1, 2, \cdots,$$

where Γ' denotes the lemniscate $|\omega(z)| = \mu_1$. It is easily shown that if by any means whatsoever we can find a series of type (1.1) which converges to $F(z)$ at every point z for which $|\omega(z)| < \mu_2$, then the coefficients of this series (which must converge uniformly, $|\omega(z)| \leqq \mu_1 < \mu_2$) are identical with the coefficients of (1.1) determined by interpolation. We shall make constant use of this remark in constructing examples, most of which will be derived from the binomial theorem.

3. **Examples.** An interesting special case is that in which $\lambda = 2$, $\alpha_1 = 1$, $\alpha_2 = -1$; the series (1.1) then has the form

$$\sum_{\nu=0}^{\infty} [a_{2\nu}(z^2-1)^{\nu} + a_{2\nu+1}(z-1)(z^2-1)^{\nu}]. \tag{3.1}$$

We shall frequently have occasion to refer to the following examples in the sequel.

(a) Let $g_1(z, q)$, q real, be an even function which, for $\Re z > 0$, coincides with a branch of the function z^{-2q} chosen so that it is regular for $\Re z > 0$ and so that $g_1(1, q) = 1$.

In particular,

$$g_1(z, \tfrac{1}{2}) = \begin{cases} 1/z, & \Re z > 0, \\ -1/z, & \Re z < 0. \end{cases}$$

We may write(7)

(6) Integrals such as this one with complex differentials are to be taken in the Lebesgue-Stieltjes sense, and those with real differentials are to be taken in the Lebesgue sense. See [16], especially pp. 64–67, for the theory of such integrals.

(7) With proper interpretation of the second member of the equation.

$$g_1(z, q) = \frac{1}{[1 + (z^2 - 1)]^q} = \sum_{\nu=0}^{\infty} C_\nu^{-q}(z^2 - 1)^\nu, \quad |z^2 - 1| < 1,$$

where $C_\nu^{-q} = (-q)(-q-1) \cdots (-q-\nu+1)/\nu! = O(n^{q-1})$, $C_0^{-q} = 1$.

(b) Let $g_2(z, q)$, q real, be an odd function which, for $\Re z > 0$, coincides with a branch of the function z^{1-2q} chosen so that it is regular for $\Re z > 0$ and so that $g_2(1, q) = 1$.

In particular,

$$g_2(z, \tfrac{1}{2}) = \begin{cases} +1, & \Re z > 0, \\ -1, & \Re z < 0. \end{cases}$$

We may write(7)

$$\begin{aligned} g_2(z, q) &= \frac{1}{[1 + (z^2 - 1)]^q} + \frac{z - 1}{[1 + (z^2 - 1)]^q} \\ &= \sum_{\nu=0}^{\infty} [C_\nu^{-q}(z^2 - 1)^\nu + C_\nu^{-q}(z - 1)(z^2 - 1)^\nu], \qquad |z^2 - 1| < 1. \end{aligned}$$

4. **Walsh's representation of $F(z)$ by columns of (1.1).** It was observed by Professor Walsh(8) that for $|z^2-1| < \mu$, the series (3.1) expresses the generating function as the sum of two components, the first of which is an even function, and the second, the product of an even function into $(z-1)$. He found that if we write $\Psi_0(z) \equiv [F(z)+F(-z)]/2 + [F(z)-F(-z)]/2z$ and $\Psi_1(z) \equiv [F(z) - F(-z)]/2z$, then $F(z) = \Psi_0(z) + (z-1)\Psi_1(z)$ and $\Psi_0(z) = \sum_0^\infty a_{2\nu}(z^2-1)^\nu$, $\Psi_1(z) = \sum_0^\infty a_{2\nu+1}(z^2-1)^\nu$, $|z^2-1| < \mu$. (Thus in §3(b), $\Psi_0 = (1+z^2-1)^{-q} = \Psi_1$.)

He further pointed out that an analogous situation exists in the general case. We may write the following equation for $|\omega(z)| < \mu_1$:

$$\begin{aligned} \Psi_K(z) &\equiv \sum_{\nu=0}^{\infty} a_{\lambda\nu+K}[\omega(z)]^\nu \\ &= \frac{1}{2\pi i}\int_{\Gamma'} \frac{F(z')}{(z' - \alpha_1) \cdots (z' - \alpha_{K+1})} \sum_0^\infty \left[\frac{\omega(z)}{\omega(z')}\right]^\nu dz' \qquad (4.1) \\ &= \frac{1}{2\pi i}\int_{\Gamma'} \frac{F(z')P_K(z')}{\omega(z') - \omega(z)} dz', \qquad K = 0, 1, \cdots, \lambda - 1, \end{aligned}$$

where $P_K(z') = \prod_{j=K+2}^{\lambda}(z' - \alpha_j)$, $K = 0, 1, \cdots, \lambda - 2$, and $P_{\lambda-1}(z') = 1$. The equalities in (4.1) are easily established by reference to (2.1) and Theorem 2.1. We then have

---

(8) The material of this section was communicated to the author in conversations with Professor Walsh.

$$(4.2)\qquad \begin{aligned} F(z) = \Psi_0(z) &+ (z-\alpha_1)\Psi_1(z) + \cdots \\ &+ (z-\alpha_1)(z-\alpha_2)\cdots(z-\alpha_{\lambda-1})\Psi_{\lambda-1}(z), \qquad |\omega(z)| < \mu_1. \end{aligned}$$

The functions $\Psi_K(z)$ are each invariant under the substitutions which leave $\omega(z)$ invariant, and in this sense are generalizations of even functions. Their Jacobi series are the second members of (4.1).

5. **Evaluation of the integrals in (4.1).** We henceforth shall assume that $\mu < \infty$. Let $w = \omega(z)/\mu$ (it is assumed invariably in the sequel that z and w are in this relationship) and let functions $z_p(w)$, $p = 1, \cdots, \lambda$, be defined by the identity

$$\omega(z) - \mu w \equiv (z - z_1(w))(z - z_2(w)) \cdots (z - z_\lambda(w)).$$

The functions $z_p(w)$ will be studied in some detail below; for the moment it suffices to observe that $\omega(z_p(w)) \equiv \mu w$, so that if $|w| < 1$ (which will be the case if z is in D) then the points $z_p(w)$ are all in D.

If we write $\omega(z') - \omega(z) = \prod_{p=1}^{\lambda}(z' - z_p(w))$ in (4.1) and notice that $\omega'(z_h(w)) = \prod_{p=1,\, p\neq h}^{\lambda}(z_h(w) - z_p(w))$, then we obtain the following equation as a simple consequence of Cauchy's integral formula:

$$(5.1)\qquad \Psi_K(z) = \sum_{p=1}^{\lambda} \frac{F(z_p(w))P_K(z_p(w))}{\omega'(z_p(w))} \equiv \Phi_K(w), \qquad K = 0, \cdots, \lambda - 1.$$

If the second members of (5.1) are replaced by their limiting values at points w for which $\omega'(z_p(w)) = 0$, $p = 1, \cdots, \lambda$, then these formulas are valid for all values of w such that $|w| < 1$.

Equation (4.2) becomes

$$(5.2)\qquad \begin{aligned} F(z_p(w)) = \Phi_0(w) &+ (z_p(w) - \alpha_1)\Phi_1(w) + \cdots \\ &+ (z_p(w) - \alpha_1)(z_p(w) - \alpha_2)\cdots(z_p(w) - \alpha_{\lambda-1})\,\Phi_{\lambda-1}(w), \qquad |w| < 1. \end{aligned}$$

We find, by combining (4.1) and (5.1), that

$$(5.3)\qquad \Phi_K(w) = \sum_{\nu=0}^{\infty} a_{\lambda\nu+K}(\mu w)^\nu, \qquad |w| < 1,\ K = 0, 1, \cdots, \lambda - 1;$$

which implies that the functions $\Phi_K(w)$ are single-valued and regular for $|w| < 1$, and that the series appearing in (5.3) are the Maclaurin series for these functions.

It is worth remarking that if $F(z_p(w)) = F(z_{p'}(w))$, for all p and p', $p = 1, \cdots, \lambda$, $p' = 1, \cdots, \lambda$, and for all w on some point set E having a limit point in the region $|w| < 1$, then $F(z) \equiv \Psi_0(z)$. For the right member of (5.2) is formally a polynomial of degree $\lambda - 1$ in the symbol $z_p(w)$, and at all but a finite number of points of E the functions $z_p(w)$ will all be distinct (they obviously fail to be so only at points corresponding to points z such

that $\omega'(z)=0$). At each point of E at which the z_p's are distinct, we have $\Phi_{\lambda-1}(w)=0$, $\Phi_{\lambda-2}(w)=0, \cdots, \Phi_1(w)=0$; and it follows from the nature of E and the regularity of the functions Φ_K that these equations must hold identically for $|w|<1$.

6. **The functions $z_p(w)$.** The functions $z_p(w)$ may be defined in such a way that they are algebraic functions([9]) of a familiar type. It seems desirable at this point to give a somewhat detailed description of a Riemann surface for the function $w=\omega(z)/\mu$ which is adapted to our needs, and of certain properties of the inverse function on this surface.

Let the distinct roots of $\omega'(z)$ be denoted by $\beta_1, \beta_2, \cdots, \beta_{\lambda_1}$, and their respective multiplicities by $m_1-1, m_2-1, \cdots, m_{\lambda_1}-1$. Let $\omega(\beta_j)/\mu=b_j$, $j=1, \cdots, \lambda_1$; these numbers b_j are not necessarily distinct. We may assume that $b_j\neq 1$, $j=1, \cdots, \lambda_1$, for such a situation can always be brought about by introducing a rotative factor into the transformation $w=\omega(z)/\mu$, which would necessitate only trivial changes in the arguments used in the sequel. We construct the Riemann surface R for this transformation as follows: The w-plane is replaced by λ superimposed planes or sheets, and the points b_j, $j=1, \cdots, \lambda_1$, and the circle γ: $|w|=1$, are plotted in each plane. We connect points b_j for which $|b_j|\geqq 1$ to ∞ by cuts which lie in the domains $|w|\geqq 1$ along rays starting at the origin and passing through these points b_j. There exists a number ρ_1, $0\leqq\rho_1<1$, such that the set of superimposed simply conected, annular-shaped regions B bounded by the curves γ, $|w|=\rho_1$, and the segment $\rho_1\leqq w\leqq 1$, have the property that none of the points b_j lie in the regions or on their boundaries except possibly on the circle γ. The points b_j for which $|b_j|<\rho_1$ are connected by a cut through all the planes, lying along a Jordan arc which passes through each of these points and through the point $w=\rho_1$, and which coincides with the positive real axis from $w=\rho_1$ to $w=\infty$, but which does not pass through any point of B.

The function $z_p(w)$ is now defined in the usual way to be single-valued and regular on the pth sheet of R except at infinity and possibly at certain of the points b_j on that sheet. The construction of R is then completed by joining the edges of the planes across the cuts properly.

If $w\neq b_j$, $j=1, \cdots, \lambda_1$, the functions $z_p(w)$ have distinct values; but if $w=b_j$, exactly m_j of these functions assume the common value β_j. All the functions are continuous for $w=b_j$, and the functions $z_p(w)$ for which $\omega'(z_p(b_j))\neq 0$ are regular for $w=b_j$. The group of m_j functions which assume the value β_j for $w=b_j$ forms a single cyclic system with respect to this value of w, because $[\partial/\partial w](\omega(z)-w)\neq 0$ [7, pp. 239–240]. For the moment let $b=b_j$, $\beta=\beta_j$, and $m=m_j$ for a fixed value of j; and let $z_{p_1}(b)=z_{p_2}(b)=\cdots=z_{p_m}(b)=\beta$. Then for w in each of a set of m regions N_h: $|w-b|\leqq\delta>0$, located on the m sheets numbered p_h, $h=1, \cdots, m$, there exists a development of the form

([9]) See [1, chaps. 2–4], and [7, pp. 233–244].

$\beta+\sum_{\nu=1}^{\infty}c_\nu[(w-b)^{1/m}]^\nu$, $c_1\neq 0$, which may be made to represent all the roots $z_{p_1}(w), \cdots, z_{p_m}(w)$, by giving to $(w-b)^{1/m}$ all of its m determinations(10). Thus letting $(w-b)_h^{1/m}$ denote the determination for z_{p_h}, we have

$$z_{p_h}(w) = \beta + \sum_1^\infty c_\nu\left[(w-b)_h^{1/m}\right]^\nu, \qquad c_1 \neq 0,\ w \in N_h. \tag{6.1}$$

Or again, we may write

$$z_{p_h}(w) - \beta = (w-b)_h^{1/m} A(W_h), \qquad w \in N_h, \tag{6.2}$$

where $W_h=(w-b)_h^{1/m}$ and $A(W_h)$ is a nonvanishing regular function of W_h for $|W_h|\leqq\delta$. We have the inequalities(11)

$$M_1 \leqq \frac{|z_{p_h}(w)-\beta|}{|w-b|^{1/m}} \leqq M_2, \qquad w\in N_h,\ w'\in N_h,\ h=1,\cdots,m, \tag{6.3}$$

$$\frac{|z_{p_h}(w)-z_{p_h}(w')|}{|w-w'|^{1/m}} \leqq M_3, \qquad w\in N_h,\ w'\in N_h,\ h=1,\cdots,m. \tag{6.4}$$

(The inequality (6.4) is easily proved by using the Heine-Borel theorem.)

The equation $z=z_p(w)$ gives a conformal one-to-one map of the annulus B in the pth sheet of R onto a simply connected region B_p in the z-plane. The region B_p lies interior to one of the curves C_k and is bounded by an arc of C_k, an analytic arc of the lemniscate Γ_{ρ_1}: $|w(z)|=\rho_1$, and analytic arcs of two orthogonal trajectories of Γ (which may coincide) which are images of the segment $\rho_1\leqq w\leqq 1$. The regions B_p, $p=1,\cdots,\lambda$, have no point in common.

The correspondence of the boundaries of B and of B_p is one-to-one and continuous. Thus the set of functions $w=z_p(e^{i\theta})$, $p=1,\cdots,\lambda$, $0\leqq\theta\leqq 2\pi$, gives a parametric representation of the lemniscate, and the arcs represented by the individual functions have no interior points in common.

Let us write $z_p(e^{i\theta})=\zeta_p(\theta)$, and arrange the subscripts of the β_j's so that the b_j's for which $|b_j|=1$ are $b_1=e^{i\xi_1}$, $b_2=e^{i\xi_2}, \cdots, b_s=e^{i\xi_s}$, where $0<\xi_1\leqq\xi_2\leqq\cdots\leqq\xi_s<2\pi$. It is obvious that we have constructed R so that the subset of the functions $z_p(w)$ assigned to the curve C_k forms a single cyclic system for $|w|=1$. Suppose that the κ_k functions $z_{p_1}, z_{p_2}, \cdots, z_{p_{\kappa_k}}$ are assigned to C_k, and that $\zeta_{p_i}(\theta)\to\zeta_{p_{i+1}}(0)$, $i=1,\cdots,\kappa_k-1$, and $\zeta_{p_{\kappa_k}}(\theta)\to\zeta_{p_1}(0)$ as $\theta\to 2\pi$, $\theta<2\pi$. We now extend the range of definition of the functions $\zeta_{p_i}(\theta)$ by means of the following conventions:

(10) [1, pp. 32–33], [7, pp. 238–240]. The reason why $c_1\neq 0$ in our case is brought out in the latter reference.

(11) Letters $M, M_1, M_2, \cdots$ will always denote finite positive constants which may depend on Γ and on $F(z)$, but which will not depend directly upon any other apparent variable or subscript, unless the contrary is implied by the use of functional notation, as in the statement of Lemma 7.1. The significance of these letters will vary with the context.

$$\zeta_{p_i}(\theta \pm 2n\pi) = \zeta_{p_{i+n}}(\theta), \qquad i = 1, \cdots, \kappa_k,\ n = 1, 2, \cdots,$$

where $p_{\nu'} = p_\nu$ for $\nu' \equiv \nu \pmod{\kappa_k}$ and $0 \leqq \theta \leqq 2\pi$. Then any one of the functions $\zeta_{p_i}(\theta)$, with its range of definition extended in this manner, will give a continuous parametric representation of the entire curve C_k, if θ is allowed to vary through $2\pi\kappa_k$ radians instead of 2π radians. That is, the function $z = \zeta_{p_i}(\theta)$ now sets up a one-to-one continuous correspondence between the points of C_k and the points of a circumference on R, considered as m-fold, closed, and of length $2\pi\kappa_k$. Any interval of values of θ of length $2\pi\kappa_k$ may be used to obtain the representation. Thus $\zeta_{p_i}(\theta)$ is now a periodic function of θ, the period being $2\pi\kappa_k$. The function is an analytic function of θ for all values of this variable except possibly at certain values which satisfy the congruences $\theta \equiv \xi_j \pmod{2\pi}$, $j = 1, \cdots, s$.

Whenever our notation implies that θ is not restricted to the interval $[0, 2\pi]$—as, for instance, in the remainder of this section and in the lemmas of §11—it is to be understood that we are using the functions $\zeta_p(\theta)$ in the extended sense.

Suppose now that for a given j we have $\zeta_{p_1}(\xi) = \zeta_{p_2}(\xi) = \cdots = \zeta_{p_{m_j}}(\xi) = \beta_j$, where $\xi \equiv \xi_j \pmod{2\pi}$. Let $\beta = \beta_j$ and $m = m_j$. For $|\theta - \xi| \leqq \pi/4$, we write $e^{i\theta} - e^{i\xi} = |2 \sin \frac{1}{2}(\theta - \xi)| e^{i\psi(\theta)}$ where

$$\psi(\theta) = \begin{cases} (3\pi + \theta + \xi)/2, & \theta \leqq \xi, \\ (\pi + \theta + \xi)/2, & \theta > \xi. \end{cases}$$

There exists a number δ', $0 < \delta' \leqq \frac{1}{4}\pi$, and a method of assigning subscripts $p_1, p_2, \cdots, p_m$ to the functions $z_p(w)$, such that, by using (6.2), we may write

$$\begin{aligned} \zeta_{p_h}(\theta) - \beta &= (e^{i\theta} - e^{i\xi})_h^{1/m} A(W_h) \\ &= |2 \sin \tfrac{1}{2}(\theta - \xi)|^{1/m} e^{i(\psi(\theta) + 2h\pi)/m} \cdot A(W_h), \qquad |\theta - \xi| \leqq \delta', \end{aligned} \tag{6.5}$$

where

$$W_h = (e^{i\theta} - e^{i\xi})_h^{1/m}$$

and $A(W_h)$ is a nonvanishing regular function of W_h for $|\theta - \xi| \leqq \delta'$. The folowing inequalities are then true for $|\theta - \xi| \leqq \delta'$, $|\theta' - \xi| \leqq \delta'$:

$$M_1 \leqq \frac{|\zeta_{p_h}(\theta) - \beta|}{|\theta - \xi|^{1/m}} \leqq M_2, \qquad h = 1, \cdots, m; \tag{6.6}$$

$$\frac{|\zeta_{p_h}(\theta) - \zeta_{p_h}(\theta')|}{|\theta - \theta'|^{1/m}} \leqq M_3, \qquad h = 1, \cdots, m. \tag{6.7}$$

An important consequence of the periodicity of the functions $\zeta_p(\theta)$ is that with each number ξ_j we may associate a positive number δ_j, $\delta_j \leqq \frac{1}{4}\pi$, with the following property: If for any p, $1 \leqq p \leqq \lambda$, we have $\zeta_p(\xi) = \beta_j$, where ξ is any

number such that $\xi \equiv \xi_j \pmod{2\pi}$, then a representation similar to (6.5) (with h suitably determined) and inequalities (6.6) and (6.7) hold for $\zeta_p(\theta)$, provided that $|\theta-\xi| \leqq \delta_j$. The δ_j is independent of p and ξ. We shall assume that $\delta_j = \delta_{j'}$ if $\xi_j = \xi_{j'}$, that $\xi_j + \delta_j < \xi_{j+1} - \delta_{j+1}$ if $\xi_j \neq \xi_{j+1}$, and that $\xi_s + \delta_s < 2\pi$, $\xi_1 - \delta_1 > 0$. It follows from these assumptions that $\zeta_p(\theta) \neq \beta_{j'}$, $j' \neq j$, for $\xi - \delta_j \leqq \theta \leqq \xi + \delta_j$, where $\xi \equiv \xi_j \pmod{2\pi}$.

Suppose now that for a given value of p and of k, the point $\zeta_p(\theta)$ lies on C_k. The length of the arc of C_k described by the point $\zeta_p(\theta)$ when θ varies from θ_1 to θ_2 is given by the formula

$$\int_{\theta_1}^{\theta_2} \left| \frac{d\zeta_p(\theta)}{d\theta} \right| d\theta = \int_{\theta_1}^{\theta_2} \left| \frac{\mu}{\omega_p(\theta)} \right| d\theta, \tag{6.8}$$

where $\omega_p(\theta) = \omega'(\zeta_p(\theta))$. It is well known that Γ is rectifiable, which implies that the integrals in (6.8) are convergent for all values of θ_1 and θ_2. We shall need the following more precise result[12]:

Lemma 6.1. $\int_\theta^{\theta+t} \mu |\omega_p(\tau)|^{-1} d\tau = O(|t|^{1/\tilde{m}})$ *uniformly for θ on any closed interval $[a, b]$, where $\tilde{m}$ is the index of the arc of Γ whose equation is $z = \zeta_p(\theta)$, $a \leqq \theta \leqq b$.*

We shall show that for each number $\xi \in [a, b]$, there exist positive numbers $M(\xi)$ and $\delta(\xi)$ independent of θ such that

$$\int_\theta^{\theta+t} \mu |\omega_p(\tau)|^{-1} d\tau \leqq M(\xi) |t|^{1/\tilde{m}}$$

for θ in the interval $I(\xi)$: $\xi - \delta(\xi) \leqq \theta \leqq \xi + \delta(\xi)$ and for $|t| \leqq \delta(\xi)$. The Heine-Borel theorem will then establish the lemma for all values of θ in the interval $[a, b]$.

The function $\omega_p(\theta) = \lambda \prod_{h=1}^{\lambda_1} (\zeta_p(\theta) - \beta_h)^{m_h - 1}$ is a continuous function of θ. If $\omega_p(\xi) \neq 0$, then $|\omega_p(\theta)|$ is bounded from zero in some neighborhood of ξ, and $\int_\theta^{\theta+t} |\omega_p(\tau)|^{-1} d\tau = O(|t|)$ uniformly for θ in some interval $I(\xi)$. On the other hand, suppose that $\zeta_p(\xi) = \beta_j$ and $\xi \equiv \xi_j \pmod{2\pi}$. Then for $|\theta - \xi| \leqq \delta_j$, $\zeta(\theta)$ satisfies an inequality similar to (6.6), and $\pi' |\zeta_p(\theta) - \beta_h|^{m_h - 1}$ is bounded from zero, where $\pi' = \prod_{h=1}^{\lambda_1}$, $h \neq j$. We have

$$\begin{aligned} \int_\theta^{\theta+t} |\omega_p(\tau)|^{-1} d\tau &\leqq M \int_\theta^{\theta+t} |\zeta_p(\tau) - \beta|^{1-m} d\tau \\ &\leqq M_1 \int_\theta^{\theta+t} |\tau - \xi|^{(1-m)/m} d\tau = O(|t|^{1/m}), \\ &\qquad |\theta - \xi| \leqq \delta_{j/2}, \ |t| \leqq \tfrac{1}{2}\delta_j, \end{aligned} \tag{6.9}$$

[12] The letters t and τ will always denote real variables. By $f(t) = O(|t|^\alpha)$, we mean that positive numbers M and T exist such that $|f(t)| \leqq M|t|^\alpha$ for $|t| \leqq T$.

where $\beta=\beta_j$ and $m=m_j$. Since $\bar{m}$ is not less than m, the proof is complete[13].

By the measure of a set on a curve C_k, we shall mean its linear Lebesgue measure. Thus the measure of a set E' on C_k which corresponds under the transformation $z=\zeta_p(\theta)$ to a set E of values of θ, is given by the formula $\int_E\mu|\omega_p(\theta)|^{-1}d\theta$; in particular, if $m(E)=0$, then $m(E')=0$ [16, pp. 123–125].

7. **The classes H^α and the existence of boundary values.** We shall say that $F(z)$ belongs to the class H_Γ^α, $\alpha>0$, or more briefly, $F(z)\in H_\Gamma^\alpha$, if[14]

$$\begin{aligned}\int_{\Gamma_\rho}|F(z)|^\alpha|dz| = \mu\sum_{p=1}^{\lambda}\int_{|w|=\rho}\frac{|F_p|^\alpha}{|\omega_p'|}|dw| = \mu\rho\int_0^{2\pi}\sum_{p=1}^{\lambda}\frac{|F_p|^\alpha}{|\omega_p'|}d\theta \\ \leqq M(F,\Gamma), \qquad 0<\rho<1,\end{aligned} \tag{7.1}$$

where Γ_ρ denotes the lemniscate $|\omega(z)|=\rho\mu$, and where $F_p=F(z_p(w))$, $\omega_p'=\omega'(z_p(w))$, $w=\rho e^{i\theta}$, and $|dz|$ and $|dw|$ denote differentials of arc length on Γ_ρ and on the circle $|w|=\rho$ respectively. The classes H_Γ^α are analogous to the well known classes of functions regular for $|w|<1$ which F. Riesz [14] called the classes H^α. We shall write H instead of H^1, and H_Γ instead of H_Γ^1; and we shall use H_Γ^∞ to denote the class of functions $F(z)$, each uniformly bounded in modulus, $z\in D$.

THEOREM 7.1. *If $F(z)\in H_\Gamma^\alpha$, then $F(z)\in H_\Gamma^\beta$, $\beta\leqq\alpha$.*

The theorem is a direct consequence of the definition of the integrals in (7.1).

THEOREM 7.2. *If $F(z)\in H_\Gamma^\alpha$, then $\Phi_K(w)\in H^{\alpha'}$, where*
(a) *$\alpha'=\alpha$ for $\bar{m}=1$, $0<\alpha\leqq\infty$,*
(b) *$\alpha'=\alpha$ for $\bar{m}\geqq1$, $\alpha\leqq1$,*
(c) *$\alpha'<\bar{m}\alpha/[\alpha(\bar{m}-1)+1]$ for $\bar{m}>1$, $1<\alpha<\infty$,*
(d) *$\alpha'<\bar{m}/(\bar{m}-1)$ for $\bar{m}>1$, $\alpha=\infty$.*

---

(13) Dr. S. E. Warschawski has pointed out in a communication to the author that the situation in Lemma 6.1 is typical of a broad class of curves. He formulates a general statement as follows:

Let the boundary of a simply connected region D contain a free arc C which consists of a finite number of arcs $\gamma_1, \gamma_2, \cdots, \gamma_n$ with bounded curvature. Let the measure of the corner at the point z_k where the two arcs γ_{k-1} and γ_k meet be π/m_k. If $z=z(w)$ maps the region $|w|<1$ conformally onto D, and if the arc $\theta_1\leqq\theta\leqq\theta_2$ on $|w|=1$ corresponds to C, then the arc length $s(\theta)$ of C satisfies the condition $s(\theta+t)-s(\theta)=O(|t|^{1/m})$, $m=\max m_k$, uniformly, $\theta_1\leqq\theta\leqq\theta_2$.

The result can be derived by reference to the results of Osgood and Taylor, these Transactions, vol. 14 (1913), p. 282, which reduce the problem to a simple computation of the type carried out in the proof of Lemma 6.1. Dr. Warschawski further observes that the condition that the γ_k's have bounded curvature can be replaced by weaker ones; e.g., the conditions required in Theorem 10 of his thesis (Mathematische Zeitschrift, vol. 35 (1932), p. 433).

Lemma 6.1 can be derived from Warschawski's result by mapping the regions B onto the unit circle.

(14) The first two integrals are to be taken in the Lebesgue-Stieltjes (or Riemann-Stieltjes) sense.

To prove the theorem, we note that the hypothesis, when taken with (7.1), implies that

$$\int_0^{2\pi}\left|\frac{F_p}{\omega_p'}\right|^{\alpha} d\theta \leqq M, \qquad \rho_1 \leqq \rho < 1, \alpha < \infty. \tag{7.2}$$

Now

$$\begin{aligned}\int_0^{2\pi} |\Phi_K(w)|^{\alpha'} d\theta &\leqq M_1 \sum_{p=1}^{\lambda}\int_0^{2\pi}\left|\frac{F_p\cdot p_K(z_p(w))}{\omega_p'}\right|^{\alpha'} d\theta \\ &\leqq M_2 \sum_{p=1}^{\lambda}\int_0^{2\pi}\left|\frac{F_p}{\omega_p'}\right|^{\alpha'} d\theta, \qquad \rho < 1.\end{aligned} \tag{7.3}$$

If $\bar{m}=1$ and $\alpha<\infty$, or in any case if $\alpha\leqq 1$, we may write the inequalities

$$\begin{aligned}\int_0^{2\pi}\left|\frac{F_p}{\omega_p'}\right|^{\alpha} d\theta &\leqq M_3\int_0^{2\pi}\frac{|F_p|^{\alpha}}{|\omega_p'|}|\omega_p'|^{1-\alpha}d\theta \\ &\leqq M_4\int_0^{2\pi}\frac{|F_p|^{\alpha}}{|\omega_p'|}d\theta \leqq M_5, \qquad \rho_1 \leqq \rho < 1,\end{aligned}$$

because if $\bar{m}=1$, $|\omega_p'|$ is uniformly bounded from zero in the regions B of R. Thus for these values of $\bar{m}$ and α, we have shown that $\int_0^{2\pi}|\Phi_K(w)|^{\alpha}d\theta\leqq M_6$ for $\rho_1\leqq\rho<1$, and a similar inequality obviously holds true for $0<\rho<1$[15]. The case $\alpha=\infty$, $\bar{m}=1$, is easily taken care of by examining the formulas (5.1).

Consider now case (*c*). Here the proof depends upon a lemma.

LEMMA 7.1. *There exists a number* $M(q)$ *such that* $\int_0^{2\pi}|\omega_p'|^{-q}d\theta\leqq M(q)$, $p=1,\cdots,\lambda$, $\rho_1\leqq\rho<1$, *where q is any number less than* $\bar{m}/(\bar{m}-1)$, $\bar{m}\geqq 1$.

The result is immediate if $\bar{m}=1$. If $\bar{m}>1$, the result is again obvious for $\rho_1\leqq\rho\leqq\rho_2<1$ if ρ_2 is chosen suitably. In this case we show that for each number ξ in the interval $0\leqq\xi\leqq 2\pi$ there exist numbers $\delta(\xi)>0$, $\rho(\xi)\geqq\rho_1$, and $M(q,\xi)=M(\xi)$ such that

$$\int_{\xi-\delta(\xi)}^{\xi+\delta(\xi)}|\omega_p'|^{-q}d\theta \leqq M(\xi)$$

for $\rho(\xi)\leqq\rho<1$. Suppose first that $\omega'(z_p(e^{i\xi}))\neq 0$. Then there exists a neighborhood of the point $w=e^{i\xi}$ in which $|\omega_p'|$ is bounded from zero, and the existence of $M(\xi)$, $\rho(\xi)$, and $\delta(\xi)$ follows at once. If $z_p(e^{i\xi})=\beta_j$, and $e^{i\xi}=b_j=b$, then according to (6.3) there is a neighborhood of the point $e^{i\xi}$ in which $|\omega_p'|^{-q}\leqq M_1(\xi)|w-b|^{-(m-1)q/m}$, where $m=m_j$. It is quite easily shown by elementary methods that any branch of the functions $(w-e^{i\xi})^{-\eta}$, $\eta<1$, which

(15) Indeed, it is well known that $\int_0^{2\pi}|\Phi_K(w)|^{\alpha}d\theta$ increases steadily with ρ [18, p. 174].

is regular for $|w|<1$, belongs to the class H[16]. The existence of $M(\xi)$, $\rho(\xi)$, and $\delta(\xi)$ is now obvious again, provided that we choose $q<m/(m-1) \leqq \bar{m}/(\bar{m}-1)$. The proof of the lemma is completed by referring to the Heine-Borel theorem and by observing that $\int_0^{2\pi}|\omega_p'|^{-q}d\theta \leqq \int_I |\omega_p'|^{-q}d\theta$, where I is any set of overlapping intervals covering the interval $[0, 2\pi]$.

To return to the proof of the theorem, we use the Hölder inequality to write

$$
(7.4)\quad \begin{aligned}
\int_0^{2\pi}\left|\frac{F_p}{\omega_p'}\right|^{\alpha'} d\theta &= \int_0^{2\pi}\left\{\frac{|F_p|^{\alpha}}{|\omega_p'|}\right\}^{\alpha'/\alpha}\cdot\left\{\left|\frac{1}{\omega_p'}\right|^{q}\right\}^{(\alpha'-\alpha'/\alpha)/q} d\theta \\
&\leqq \left\{\int_0^{2\pi}\frac{|F_p|^{\alpha}}{|\omega_p'|}d\theta\right\}^{\alpha'/\alpha}\left\{\int_0^{2\pi}\left|\frac{1}{\omega_p'}\right|^{q}d\theta\right\}^{(\alpha'-\alpha'/\alpha)/q}
\end{aligned}
$$

for $\rho_1 \leqq \rho < 1$. In this inequality, we must have $\alpha' < \alpha$ and $(\alpha'/\alpha)+(\alpha'-\alpha'/\alpha)/q = 1$, or $\alpha' = q\alpha/(q+\alpha-1)$. By referring to (7.2) and the lemma, we see that the third member of (7.4) is bounded for $\rho_1 \leqq \rho < 1$ if $q < \bar{m}/(\bar{m}-1)$; that is, if $\alpha' < \bar{m}\alpha/[\alpha(\bar{m}-1)+1]$. The proof of this section of the theorem is then completed by using (7.3).

Case (d) follows immediately from (5.2) and the lemma. The proof of the theorem is now complete.

It is easily shown that the inequality in (d) cannot be replaced by an equality. An example is given by the function $g_2(z, \frac{1}{2})$, which belongs to the class H_Γ^α, where Γ is now the lemniscate $|z^2-1|=1$, and for which $\Phi_0(w)$ and and $\Phi_1(w)$ obviously belong to the classes $H^{\alpha'}$, $\alpha' < 2$, but not to the class H^2.

Similarly, the inequality in (c) cannot be replaced by an equality. An example to prove this may be constructed as follows: Let $F(z) \equiv 0$ for $\Re z < 0$, and let $F(z)$ denote for $|z^2-1|<1$, $\Re z > 0$, a branch of the function $z^{-1/2} \cdot [\log(e/z^2)]^{-3/4}$ regular in the simply connected region defined by these inequalities. Then $F(z) \in H_\Gamma^2$, where Γ is again the lemniscate $|z^2-1|=1$; and $\Phi_1(w) \in H^{\alpha'}$, $\alpha' < 4/3$, but $\Phi_1(w)$ does not belong to the class $H^{4/3}$. The reader will have no difficulty in supplying a proof.

When $F(z) \in H_\Gamma^\alpha$, it does not follow in general that $\Psi_K(z) \in H_\Gamma^\alpha$ (an example is again given by $g_2(z, \frac{1}{2})$), but the following result is an immediate consequence of Theorem 7.2.

THEOREM 7.3. *If* $F(z) \in H^\alpha$, *then*

$$
\int_{\Gamma_\rho} |\Psi_K(z)|^{\alpha'} |\omega'(z)| \, |dz| < M, \qquad K = 0, 1, \cdots, \lambda - 1,
$$

for $0 < \rho < 1$, *where* α' *is given by the formulas in Theorem* 7.2.

---

(16) We have $\int_0^{2\pi}|1+\rho^2-2\rho\cos(\theta-\xi)|^{-q/2}d\theta \leqq \int_0^{2\pi}(2\rho)^{-q/2}|1-\cos(\theta-\xi)|^{-q/2}d\theta < M(\eta)$, $\frac{1}{2} \leqq \rho < 1$.

The next two theorems are in the nature of converses to Theorems 7.2 and 7.3.

THEOREM 7.4. *If* $\Phi_K(w) \in H^\alpha$, $K=0, 1, \cdots, \lambda-1$, *then*

$$\int_{\Gamma_\rho} |F(z)|^\alpha |\omega'(z)| |dz| < M$$

for $0<\rho<1$.

THEOREM 7.5. *If* $\Psi_K(z) \in H_\Gamma^\alpha$, $K=0, 1, \cdots, \lambda-1$, *then* $F(z) \in H_\Gamma^\alpha$.

The proofs are easily supplied by referring to (4.2) and (5.2). The example $g_1(z, \frac{1}{2})$ shows that in general the factor $|\omega'(z)|$ cannot be omitted from the integrand in Theorem 7.4.

We now consider the existence of boundary values. We define an S-path(17) of a Jordan region T bounded by a contour C to be an analytic Jordan arc terminating at a point Z of C and lying in a closed triangular subregion of the closed region $\overline{T}$, the boundary of which has a vertex at Z but has no other point in common with C, and is at no point tangent to C.

THEOREM 7.6. *If* $F(z) \in H_\Gamma^\alpha$, *there exist finite-valued functions* $\Phi_K^*(w)$ *and* $F^*(z)$ *such that*

(a) $\Phi_K(w) \to \Phi_K^*(W)$, $K=0, 1, \cdots, \lambda-1$, *for almost every point* W *for which* $|W|=1$, *as* $w \to W$ *along any S-path of the regions* $|w|<1$;

(b) $F(z) \to F^*(Z)$ *for almost every point* Z *on* Γ *as* $z \to Z$ *along any S-path of* D(18);

(c) $F^*(z_p(w)) = \Phi_0^*(w) + (z_p(w)-\alpha_1)\Phi_1^*(w) + \cdots + (z_p(w)-\alpha_1)(z_p(w)-\alpha_2) \cdots (z_p(w)-\alpha_{\lambda-1})\Phi_{\lambda-1}^*(w)$ *for almost every point* w *on the circle* γ *in the pth sheet of* R;

(d) *for almost every* w *for which* $|w|=1$

$$\Phi_K^*(w) = \sum_{p=1}^{\lambda} \frac{F^*(z_p(w))P_K(z_p(w))}{\omega'(z_p(w))}, \qquad K = 0, 1, \cdots, \lambda - 1;$$

(e) *for* α' *given by the formulas in Theorem* 7.2,

$$\int_\Gamma |F^*(z)|^\alpha |dz| < \infty, \quad \int_\gamma |\Phi_K^*(w)|^{\alpha'} |dw| < \infty, \ K = 0, 1, \cdots, \lambda - 1.$$

Part (a) is a consequence of Theorem 7.2 and a theorem of F. Riesz [14](19).

(17) S for Stolz.

(18) The implication here is that $F^*(z)$ may be m-valued at a multiple point of order m, but is single-valued at all ordinary points of Γ.

(19) See also [22, p. 162].

We temporarily define $F^*(z_p(w)$ for $|w|=1$ by the second member of the equality in (c). Now if the point $z_p(w)$ traverses an S-path in B_p terminating at $z_p(W)$ on Γ, then the point w traverses an S-path of the region B on the pth sheet of R, terminating at the point W on that sheet. (This is an immediate consequence of the mapping properties of the analytic function $w=\omega(z)/\mu$.) The set of points on the circle γ in the pth sheet of R at which one or more of the limits in (a) fail to exist is of measure zero, and (as observed at the end of §6) the corresponding set σ on Γ is also of measure zero. It is now easily seen from (5.2) that the statement in (b) is true for all points Z of Γ which are not in σ, which are also boundary points of B_p, and for which $\omega(Z)/\mu\neq 1$. For convenience we may add the set of points Z defined by the equation $\omega(Z)/\mu=1$ to the set of excepted points in (b), and since the reasoning of this paragraph applied to S-paths in all of the regions B_p, $p=1, \cdots, \lambda$, the proof of (b) is complete.

Part (d) is now a consequence of (5.1).

Part (e) is proved immediately by referring to (7.1), parts (a) and (b) of the present theorem, and the lemma of Fatou [16, p. 29]. The proof of Theorem 7.6 is complete.

Now that the existence of boundary values and the validity of the limit in part (b) of Theorem 7.6 has been established, we shall revise the definition of $F^*(z)$ as follows: On each curve C_k, $F^*(Z)$ shall be the unique limit approached by the given function $F(z)$ as $z\rightarrow Z$ along any S-path of D_k, at each point Z of C_k for which this limit exists and is finite. At all other points of C_k, we let $F^*(z)=0$. The function $F^*(z)$ so defined exists everywhere on Γ, is single-valued except at the multiple points, and coincides with the previously defined boundary value function wherever the latter exists.

It will be convenient henceforth to use the single symbol $F(z)$ to designate the complete function consisting of $F(z)$, $|\omega(z)|<\mu$, and $F^*(z)$, $|\omega(z)|=\mu$; and we shall use the symbols $\Phi_K(w)$, $K=0, 1, \cdots, \lambda-1$, in a similarly extended sense.

The case $\alpha=1$ is of especial importance because of this theorem:

THEOREM 7.7. *If $F(z)\in H_\Gamma$, then the series $\sum_{\nu=0}^{\infty}a_{\lambda\nu+K}\mu^\nu e^{i\nu\theta}$, $K=0, 1, \cdots, \lambda-1$, are the Fourier series respectively of the functions $\Phi_K(e^{i\theta})$, $K=0, 1, \cdots, \lambda-1$.*

This result is a consequence of Theorem 7.2, equation (5.3), and a theorem of F. and M. Riesz [15, p. 42], which states that our conclusion is a necessary and sufficient condition for $\Phi_K(w)\in H$.

Conversely, if it be known of a series of the form (1.1) that for some number μ the series $\sum a_{\lambda\nu+K}\mu^\nu e^{i\nu\theta}$, $K=0, 1, \cdots, \lambda-1$ are all Fourier series, then there exist functions $\Phi_K(w)\in H$, $K=0, 1, \cdots, \lambda-1$ defined by equation (5.3), with boundary values on the unit circle for which these series are respectively the Fourier series; and there exists a function $F(z)$ defined by equation (4.2),

such that $F(z)\omega'(z)\in H_\Gamma$ (where Γ is the lemniscate $|\omega(z)|=\mu$) and for which the original series is the Jacobi series with respect to the points α_j[20].

Henceforth we shall assume that $F(z)\in H_\Gamma$, unless a statement is made to the contrary. We shall also make these abbreviations: $F(z_p(e^{i\theta}))=f_p(\theta)$, $\Phi_K(e^{i\theta})=\phi_K(\theta)$, $P_K(z_p(e^{i\theta}))=p_{K,p}(\theta)$.

Theorem 7.7 and Theorem 7.6, parts (d) and (e) enable us to write[21]

$$
\begin{aligned}
\mu^\nu a_{\lambda\nu+K} &= \frac{1}{2\pi}\int_0^{2\pi}\phi_K(\theta)e^{-i\nu\theta}d\theta \\
&= \frac{1}{2\pi}\int_0^{2\pi}e^{-i\nu\theta}\sum_{p=1}^{\lambda}\left(\frac{f_p(\theta)p_{K,p}(\theta)}{\omega_p(\theta)}\right)d\theta \\
&= \frac{1}{2\pi}\sum_{p=1}^{\lambda}\int_0^{2\pi}\frac{f_p(\theta)p_{K,p}(\theta)}{\omega_p(\theta)}e^{-i\nu\theta}d\theta \\
&= \frac{\mu^\nu}{2\pi i}\int_\Gamma\frac{F(z)dz}{(z-\alpha_1)\cdots(z-\alpha_{K+1})[\omega(z)]^\nu}, \quad K=0,1,\cdots,\lambda-1.
\end{aligned} \tag{7.5}
$$

Equation (7.5) is valid for all integral values of ν if the first member is replaced by zero when ν is negative. In particular, if $\nu=-1$ and $K=\lambda-1$, we have

$$
\frac{1}{2\pi i\mu}\int_\Gamma F(z)dz=\frac{1}{2\pi}\int_0^{2\pi}\phi_{\lambda-1}(\theta)e^{i\theta}d\theta=\frac{1}{2\pi i}\int_\gamma\Phi(w)dw=0 \tag{7.6}
$$

[15, p. 42]. A function $F_k(z)$ identical with $F(z)$ in $\overline{D}_k$ and identically zero elsewhere, will also be of class H_Γ. We may accordingly apply to $F_k(z)$ the argument which led to (7.6), and we find that $\int_\Gamma F_k(z)dz=\int_{C_k}F(z)dz=0$. Also, by Cauchy's integral theorem, $\int_{C_k}S_n(z)dz=0$. Combining these remarks, we have

THEOREM 7.8. $\int_{C_k}F(z)dz=\lim_{n\to\infty}\int_{C_k}S_n(z)dz=0$, $k=1,\cdots,\lambda'$.

8. **The coefficients.** We now present a few typical results concerning the order of magnitude of the quantities $\mu^\nu|a_{\lambda\nu+K}|$, $K=0,\cdots,\lambda-1$, $\nu=0,1,2,\cdots$.

Our first theorem follows immediately from (7.5).

THEOREM 8.1. *If* $|F(z)|\leq M(F)$, $z\in\Gamma$, *then* $\mu^\nu|a_{\lambda\nu+K}|\leq M(F)M_1(\Gamma)$, $K=0,1,\cdots,\lambda-1$, $\nu=0,1,2,\cdots$, *where* $M_1(\Gamma)$ *is independent of* $F(z)$.

---

[20] Strictly speaking, we are using the symbol H_Γ in a slightly extended sense here, because in the present instance the lemniscate Γ need not be the lemniscate of convergence of the Jacobi series.

[21] See [16, pp. 36–38], for the theorem on change of variables in the Lebesgue-Stieltjes integral.

Our second theorem is simply the Riemann-Lebesgue theorem [22, p. 18] applied to the Fourier coefficients of the functions $\phi_K(\theta)$:

THEOREM 8.2. $\lim_{\nu\to\infty} \mu^\nu a_{\lambda\nu+K}=0$, $K=0, 1, \cdots, \lambda-1$.

Suppose now that $|F(z)|$ is uniformly bounded on Γ. Let $C_p(\delta)$, $\delta>0$, denote the modulus of continuity [22, p. 17] of the periodic function $f_p(\theta)$, and let $C(\delta)$ denote for each δ the largest of the numbers $C_p(\delta)$, $p=1, \cdots, \lambda$.

THEOREM 8.3. $|\mu^\nu a_{\lambda\nu+K}| < M[C(\pi/\nu)+(\pi/\nu)^{1/\bar{m}}]$, $K=0, 1, \cdots, \lambda-1$, $\nu=1, 2, \cdots$.

For the proof, we write

$$2\pi\mu^\nu a_{\lambda\nu+K} = \frac{1}{2}\int_0^{2\pi} [\phi_K(\theta) - \phi_K(\theta + \pi/\nu)]e^{-i\nu\theta}d\theta,$$

and by reference to (7.5) we see that it is sufficient to show that the integrals

$$I_{K,p} = \int_0^{2\pi} \left| \frac{f_p(\theta)}{\omega_p(\theta)} p_{K,p}(\theta) - \frac{f_p(\theta + \pi/\nu)}{\omega_p(\theta + \pi/\nu)} p_{K,p}(\theta + \pi/\nu) \right| d\theta,$$
$$K = 0, 1, \cdots, \lambda - 1, \; p = 1, \cdots, \lambda,$$

each satisfy an inequality similar to the one in the statement of the theorem. We have

$$\begin{aligned} I_{K,p} &\leqq \int_0^{2\pi} \left| \frac{p_{K,p}(\theta)}{\omega_p(\theta)} \right| | f_p(\theta) - f_p(\theta + \pi/\nu) | \, d\theta \\ &\quad + \int_0^{2\pi} | f_p(\theta + \pi/\nu) | \left| \frac{p_{K,p}(\theta)}{\omega_p(\theta)} - \frac{p_{K,p}(\theta + \pi/\nu)}{\omega_p(\theta + \pi/\nu)} \right| d\theta \\ &= J_1 + J_2. \end{aligned}$$

It is obvious that $J_1 \leqq M_1 C(\pi/\nu)$. We shall have shown that $J_2 = O[(\pi/\nu)^{1/\bar{m}}]$ if we establish that

$$\text{(8.1)} \qquad \int_0^{2\pi} \left| \frac{p_{K,p}(\theta)}{\omega_p(\theta)} - \frac{p_{K,p}(\theta + t)}{\omega_p(\theta + t)} \right| d\theta = O(|t|^{1/\bar{m}}),$$
$$K = 0, 1, \cdots, \lambda - 1, \; p = 1, \cdots, \lambda;$$

and the remainder of our proof will be concerned with this relation.

Assuming that $0 \leqq t \leqq \delta_j/2$, $j=1, \cdots, s$, we write

$$\text{(8.2)} \qquad \begin{aligned} \int_0^{2\pi} &\leqq \sum_{j=1}^{s}\left(\int_{\xi_j-\delta_j}^{\xi_j-t} + \int_{\xi_j}^{\xi_j+\delta_j-t}\right) + \sum_{j=1}^{s} \left(\int_{\xi_j-t}^{\xi_j} + \int_{\xi_j+\delta_j-t}^{\xi_j+\delta_j}\right) \\ &\quad + \sum{}' \int_{\xi_j+\delta_j}^{\xi_{j+1}-\delta_{j+1}} = \sum_{j=1}^{s} \sigma_j + \Sigma_2 + \Sigma_3, \end{aligned}$$

where $\xi_0+\delta_0=0$, $\xi_{s+1}-\delta_{s+1}=2\pi$, and where in $\sum'$ the index j assumes all values from zero to s for which $\xi_j<\xi_{j+1}$. The function $p_{K,p}(\theta)/\omega_p(\theta)$ is an analytic function of θ for $\xi_j+\delta_j\leqq\theta\leqq\xi_{j+1}-\delta_{j+1}+\frac{1}{2}\delta_{j+1}$, $j=0, 1, \cdots, s-1$, $\xi_j<\xi_{j+1}$, and for $\xi_s+\delta_s\leqq\theta\leqq 2\pi+\delta_1$; whence $\Sigma_3=O(t)$. By Lemma 6.1, $\Sigma_2=O(t^{1/\bar{m}})$. If $\omega_p(\xi_j)\neq 0$, then $\sigma_j=O(t)$. If on the contrary, $\zeta_p(\xi_j)=\beta_j$, then it is rather easily shown by using (6.5) and the identity $AB-CD=A(B-D)+D(A-C)$ that

$$\sigma_j \leqq M\left(\int_{\xi_j-\delta_j}^{\xi_j-t}+\int_{\xi_j}^{\xi_j+\delta_j-t}\right)\left|\,\left|\frac{1}{2\sin\frac{1}{2}(\theta-\xi_j)}\right|^{(m_j-1)/m_j}\right.$$

$$\left. -\left|\frac{1}{2\sin\frac{1}{2}(\theta+t-\xi_j)}\right|^{(m_j-1)/m_j}\right|d\theta+O(t^{-1/m_j}) = O(t^{-1/\bar{m}}),$$

which establishes (8.1) for $t\geqq 0$. The proof of the theorem is now complete.

We note in passing that (8.1) may easily be established for negative values of t by a slight modification of the above argument.

In certain senses, the estimate of the coefficients given by Theorem 8.3 cannot be improved. For example, the function $g_2(z;\frac{1}{2})$ (for which the lemniscate of convergence has the index 2) has the property that $C(\delta)\equiv 0$, but the coefficients of its Jacobi series are $O(n^{-1/2})$, but not $o(n^{-1/2})$. Again, by using certain examples([22]) in the theory of trigonometric series, it is easy to construct functions of class H_Γ, with $\bar{m}=1$, for which $C(\delta)\leqq M\delta^\eta$, $0<\eta<1$, and for which the quantities $\mu^\nu|a_{\lambda_\nu+K}|$ are $O(n^{-\eta})$ but not $o(n^{-\eta})$.

THEOREM 8.4. *If $F(z)$ is of bounded variation on Γ, then*
(a) $\mu^\nu a_{\lambda_\nu+K}=O(\nu^{-1/\bar{m}})$, $\bar{m}>1$,
(b) $\mu^\nu a_{\lambda_\nu+K}=o(\nu^{-1})$, $\bar{m}=1$.

It follows at once from the continuous, one-to-one nature of the correspondence between the boundaries of the regions B and B_p, that under the present hypothesis on $F(z)$, the functions $f_p(\theta)$ are of bounded variation on any finite interval, as are also the functions $\omega_p(\theta)$ and $p_{K,p}(\theta)$. If $\bar{m}=1$, it is a consequence of the formulas in Theorem 7.6 (d) that the functions $\phi_K(\theta)$ are of bounded variation. Therefore since they are the boundary values of functions of class H, they are absolutely continuous [15; 22, p. 158], and their Fourier coefficients are $o(n^{-1})$ [22, p. 18]. This proves part (b) of the theorem.

Turning to part (a), we write

$$\int_0^{2\pi}\phi_K(\theta)e^{-i\nu\theta}d\theta=\sum{}'\int_{\xi_j-\delta_j}^{\xi_j+\delta_j}\phi_K(\theta)e^{-i\nu\theta}d\theta+\sum{}'\int_{\xi_j+\delta_j}^{\xi_{j+1}-\delta_{j+1}}\phi_K(\theta)e^{-i\nu\theta}d\theta,$$

$$=\Sigma_1+\Sigma_2,$$

where $\xi_0\pm\delta_0=0$, $\xi_{s+1}-\delta_{s+1}=2\pi$, and $\sum'$ has the same significance as in

([22]) See for instance [22, p. 38, ex. 3] (the conjugate is $\in$ lip α, by the theorem of Privaloff [22, p. 156]).

(8.2). The function $\phi_K(\theta)$ is of bounded variation, $\xi_j+\delta_j\leqq\theta\leqq\xi_{j+1}-\delta_{j+1}$, $j=0, 1, \cdots, s$, $\xi_j<\xi_{j+1}$, so by a classical theorem [22, p. 18], $\Sigma_2=O(\nu^{-1})$. If $\omega_p(\xi_j)\neq 0$, then $f_p(\theta)p_{K,p}(\theta)/\omega_p(\theta)$ is of bounded variation for θ in the closed interval I: $[\xi_j-\delta_j, \xi_j+\delta_j]$, and $\int_I[f_p(\theta)p_{K,p}(\theta)e^{-i\nu\theta}/\omega_p(\theta)]d\theta=O(\nu^{-1})$, by the classical theorem. Suppose now that $\zeta_{p_1}(\xi)=\zeta_{p_2}(\xi)=\cdots=\zeta_{p_m}(\xi)=\beta$, where $\xi=\xi_j$. Then by referring to (6.5) we may write, for $\theta\in I$,

$$\sum_{h=1}^{m}\frac{f_{p_h}(\theta)p_{K,p_h}(\theta)}{\omega_{p_h}(\theta)}=\sum_{h=1}^{m}\frac{f_{p_h}(\theta)\chi_{K,p_h}(\theta)}{(\zeta_{p_h}(\theta)-\beta)^{m-1}}$$

$$(8.3)\qquad =\frac{1}{|\theta-\xi|^{(m-1)/m}}\left[\frac{e^{-i[(m-1)/m]\psi(\theta)}\,|\theta-\xi|^{(m-1)/m}}{|2\sin\frac{1}{2}(\theta-\xi)|^{(m-1)/m}}\cdot\sum_{h=1}^{m}\frac{f_{p_h}(\theta)\chi_{K,p_h}(\theta)e^{2\pi ih/m}}{[A(W_h)]^{m-1}}\right],$$

where $\chi_{K,p_h}(\theta)$ is a rational function of $\zeta_{p_h}(\theta)$ with no poles for $\theta\in I$. Since $\chi_{K,p_h}(\theta)/[A(W_h)]^{m-1}$ is an absolutely continuous function of θ for $\theta\in I$, and $\psi(\theta)$ is a step function, and $|\theta-\xi|/|\sin\frac{1}{2}(\theta-\xi)|$ is an analytic function of θ for $\theta\in I$, it follows that the quantity in brackets in (8.3) is of bounded variation for $\theta\in I$. Let us denote this quantity by $X(\theta)$. Another classical theorem [3, p. 494] now states that

$$\lim_{n\to\infty}\nu^{-1/m}\int_I X(\theta)\frac{e^{-i\nu\theta}}{|\theta-\xi|^{(m-1)/m}}d\theta=K_1X(\xi+0)+K_2X(\xi-0),$$

where K_1 and K_2 are complex constants depending on m. The proof of part (a) is complete.

We observe that

$$(8.4)\qquad X(\xi\pm 0)=e^{-i[(m-1)/m]\psi(\xi\pm 0)}\chi_{K,p_1}(\xi)\cdot c_1^{1-m}\sum_{h=1}^{m}f_{p_h}(\xi\pm 0)e^{2\pi ih/m}.$$

If $F(z)$ is of bounded variation on Γ, it can have only simple discontinuities on Γ. It can be shown that such discontinuities cannot occur on the curves C_k, which means that we can replace $f_{p_h}(\xi\pm 0)$ by $f_{p_h}(\xi)$ in (8.4). If we further suppose that $f_{p_1}(\xi)=f_{p_2}(\xi)=\cdots=f_{p_m}(\xi)$, then obviously $X(\xi\pm 0)=0$. Thus we have the following theorem:

THEOREM 8.5. *If $F(z)$ is single-valued and of bounded variation on Γ, then* $\mu^\nu a_{\lambda\nu+K}=o(\nu^{-1/\bar m})$, $\bar m\geqq 1$.

The function $g_2(z;\frac{1}{2})$, which is absolutely continuous on each of the two contours of its lemniscate of convergence, has coefficients which are $O(n^{-1/\bar m})$, but not $o(n^{-1/\bar m})$; so the estimate in Theorem 8.4(a) cannot be improved. We cannot replace $o(\nu^{-1/\bar m})$ in Theorem 8.5 by $O(\nu^{-(1/\bar m)-\epsilon})$, where ϵ is any fixed

positive number, because the function $g_2(z; q)$, $q<\frac{1}{2}$, is absolutely continuous on its entire lemniscate of convergence (for which $\bar{m}=2$), but its Jacobi coefficients are $O(n^{q-1})$ but not $o(n^{q-1})$.

We conclude this section with two results analogous respectively to Parseval's theorem and to the Riesz-Fisher theorem.

Let Q_p denote the series $\sum_{\nu=0}^{\infty}\sum_{K=0}^{\lambda-1}|\mu^\nu a_{\lambda\nu+K}|^p$.

THEOREM 8.6. *If $F(z)\in H_\Gamma^\alpha$, $1<\alpha\leqq 2$, then Q_β is convergent, where*
(a) *$\beta=\alpha/(\alpha-1)$ for $\bar{m}=1$,*
(b) *$\beta>\alpha\bar{m}/(\alpha-1)$ for $\bar{m}>1$.*

The theorem is an immediate consequence of Theorem 7.2, parts (a) and (c), and of the Hausdorff-Young theorem [22, p. 190]. The class H^2 is of particular interest because in this case $\beta=\alpha$; but the class H_Γ^2 is of comparable interest only when $\bar{m}=1$. The example $g(z; \frac{1}{2})$ is worth mentioning in this connection. This function is of class H_Γ^∞, and the nth partial sum of the corresponding series Q_2 is twice the Landau upper bound for the modulus of the nth partial sum of the Taylor series for an arbitrary function of class H^∞, and is asymptotic to $(2 \log n)/\pi$.

THEOREM 8.7. *Let $a_0, a_1, \cdots$ be any sequence of numbers such that for some $\mu>0$, $\sum_{\nu=0}^{\infty}\sum_{K=0}^{\lambda-1}|\mu^\nu a_{\lambda\nu+K}|^\alpha$ is finite, $1<\alpha\leqq 2$. There exists a function $F(z)$ regular and single-valued for $|\omega(z)|<\mu$ such that $\int_{\Gamma_\rho}|F(z)|^\beta|\omega'(z)||dz|<M$ for $0<\rho<1$, where $\beta=\alpha/(\alpha-1)$, and for which the numbers a_n are the coefficients of the Jacobi series with respect to the points α_i.*

It is easily established by means of the Hausdorff-Young theorem [33, p. 190], and with the aid of the theory of Abel means [33, p. 87], that functions $\Phi_K(w)\in H^\beta$ exist for which the numbers $\mu^\nu a_{\lambda\nu+K}$ are the Maclaurin coefficients; and the remainder of the proof may be supplied by reference to Theorem 7.4 and the remark at the end of §2.

9. **Convergence theorems obtained by using the functions $\phi_K(\theta)$.** The work of §§4–7 allows us to answer many questions concerning the convergence and summability of the Jacobi series by merely referring directly to the theory of Fourier series. We shall of course not attempt to give a catalogue of such convergence theorems here, but shall refer briefly to certain results which can perhaps be considered typical.

The first of these is an analogue of the Fejér-Lebesgue-Hardy theorem [22, p. 49].

THEOREM 9.1. $\text{Lim}_{n\to\infty} S_n^{(r)}(z)=F(z)$, *$r>0$, almost everywhere on Γ.*

The theorems of §7 and the Fejér-Lebesgue-Hardy theorem establish this result immediately for the functions $\Psi_K(z)$ and their Jacobi series; and the proof may be completed by using a theorem of the author on the Cesàro method of summation [4, pp. 707–708], and Theorem 7.6(c).

We turn next to convergence in the mean. It is well known that if $|\phi_K(\theta)|^p$, $p \geqq 1$, is integrable for $0 \leqq \theta \leqq 2\pi$, then the Fourier series for $\phi_K(\theta)$ converges in the mean to $\phi_K(\theta)$ with index p', where $p' = p$ if $p > 1$, $0 < p' < 1$ if $p = 1$ [22, p. 153]. The proof of the following theorem is easily supplied by using the results of §7 and the Minkowski inequality [18, p. 384].

THEOREM 9.2. *If $F(z) \in H^{\alpha}_{\Gamma}$, $\alpha \geqq 1$, then*

$$\text{(a)} \quad \lim_{n\to\infty} \int_\Gamma \left| \Psi_K(z) - \sum_{\nu=0}^{n} a_{\lambda\nu+K}[\omega(z)]^\nu \right|^{\alpha'} |\omega'(z)| \, |dz| = 0, K = 0, 1, \cdots, \lambda - 1,$$

$$\text{(b)} \quad \lim_{n\to\infty} \int_\Gamma |F(z) - S_n(z)|^{\alpha'} |\omega'(z)| \, |dz| = 0.$$

The number α' is described by the formulas in Theorem 7.2 if $\alpha > 1$. If $\alpha = 1$, then $0 < \alpha' < 1$.

Uniform convergence theorems may be obtained by imposing suitable conditions directly on the functions $\phi_K(\theta)$. We have, for instance, as an analogue of Dirichlet's theorem [22, p. 25] the following result:

THEOREM 9.3. *If the functions $\phi_K(\theta)$, $K = 0, 1, \cdots, \lambda - 1$, are of bounded variation for $0 \leqq \theta \leqq 2\pi$, then $\lim_{n\to\infty} S_n(z) = F(z)$ uniformly for z in $\overline{D}$.*

Obviously there is a theorem of this type for each of the many tests for convergence of the Fourier series. But it would seem to be of interest to derive such results from conditions imposed directly on the function $F(z)$ rather than on the functions $\phi_K(\theta)$, particularly if $\bar{m} > 1$. The functions $\Psi_K(z)$ may behave much less "smoothly" than $F(z)$ in the neighborhood of a multiple point on Γ, and the convergence properties of the series $\sum_0^\infty a_{\lambda\nu+K}[\omega(z)]^\nu$ may not adequately reflect those of the Jacobi series for the function $F(z)$. For example, the function $g_2(z, q)$, $0 < q < \frac{1}{2}$, is absolutely continuous on its lemniscate of convergence, and its Jacobi series converges uniformly on this lemniscate[23]; but the corresponding functions $\Psi_0(z)$ and $\Psi_1(z)$ are both infinite for $z = 0$ and their Jacobi series are properly divergent at this point.

In the remainder of this paper we shall introduce methods which enable us to derive convergence theorems without reference to the properties of the functions $\Psi_K(z)$ and their series.

10. **The function $H(Z, w)$ and the integral formulas for $S_{\lambda n+\lambda-1}(z; F)$.** Let $F(z)$ be any analytic function, not necessarily of class H_Γ, for which Γ is the lemniscate of convergence. We define the function $H(Z, w; F)$, or $H(Z, w)$, for any Z and for $|w| < 1$, as follows:

$$(10.1) \quad H(Z, w) = \Phi_0(w) + (Z - \alpha_1)\Phi_1(w) + \cdots + (Z - \alpha_1) \cdots (Z - \alpha_{\lambda-1})\Phi_{\lambda-1}(w).$$

[23] See Theorem 11.7.

The variables Z and w are to be considered as independent here. For a fixed w, this function is a polynomial in Z which coincides with $F(Z)$ in the points $Z=z_p(w)$, $p=1, \cdots, \lambda$. Therefore for $w\neq b_j$, $j=1, \cdots, \lambda_1$, we may use the Lagrange interpolation formula [20, p. 50] to write([24])

$$H(Z, w) = \sum_{p=1}^{\lambda} \frac{F(z_p(w))}{\omega'(z_p(w))} \frac{\omega(Z) - w}{Z - z_p(w)}, \qquad |w| < 1. \tag{10.2}$$

The quotients in this formula are of course supposed to be defined by their limiting values for $Z=z_p(w)$, $p=1, \cdots, \lambda$. If $F(z)\equiv 1$, then $H(Z, w)\equiv 1$, so for any number f we have

$$H(Z, w) - f = \sum_{p=1}^{\lambda} \frac{F(z_p(w)) - f}{\omega'(z_p(w))} \frac{\omega(Z) - \mu w}{Z - z_p(w)}, \tag{10.3}$$
$$|w| < 1, \ w \neq b_j, \ j = 1, \cdots, \lambda_1.$$

THEOREM 10.1. *If $F(z)\in H_\Gamma^\alpha$, then $H(Z, w)$, considered as a function of w, belongs to the class $H^{\alpha'}$, where α' is given by the formulas of Theorem 7.2.*

The result is an immediate consequence of (10.1) and Theorem 7.2.

THEOREM 10.2. *If $F(z)\in H_\Gamma^\alpha$, there exists a finite-valued function $H^*(Z, w)$ defined for each value of Z and for almost every w, $|w|=1$, by the equation*

$$H^*(Z, w) = \Phi_0(w) + (Z - \alpha_1)\Phi_1(w) + \cdots + (Z - \alpha_1) \cdots (Z - \alpha_{\lambda-1})\Phi_{\lambda-1}(w),$$

and such that

(a) *$H(Z, w)\to H^*(Z, W)$ and $H(z_p(w), w)\to H^*(z_p(W), W)$, $p=1, \cdots, \lambda$, for almost every point W for which $|W|=1$ as $w\to W$ along any S-path of the regions $|w|<1$;*

(b) $$H^*(Z, w) = \sum_{p=1}^{\lambda} \frac{F(z_p(w))}{\omega'(z_p(w))} \frac{\omega(Z) - \mu w}{Z - z_p(w)}$$

for almost every w for which $|w|=1$;

(c) $$\int_\gamma |H(Z, w)|^{\alpha'} |dw| < \infty, \qquad \int_\gamma |H(z_p(w), w)|^{\alpha} |dw| < \infty,$$
$$p = 1, \cdots, \lambda,$$

where α' is given by the formulas in Theorem 7.2.

The theorem is easily proved by using (10.1), (10.2), and Theorem 7.6. We shall henceforth use the single symbol $H(Z, w)$ to designate the complete function consisting of $H(Z, w)$, $|w|<1$, and $H^*(Z, w)$, $|w|=1$. Let $h_p(\theta', \theta)$

([24]) The function $H(Z, w)$ is independent of the order of the points α_j. Thus in studying this function, we are turning to Jacobi's original point of view.

$=H(z_p(e^{i\theta'}), e^{i\theta})$ and let

$$(10.4)\quad \chi_{p,q}(\theta', \theta) = \frac{1}{\omega_q(\theta)} \frac{\mu e^{i\theta'} - \mu e^{i\theta}}{\zeta_p(\theta') - \zeta_q(\theta)} = \frac{\prod_{j=1, j\neq p}^{\lambda} (\zeta_q(\theta) - \zeta_j(\theta'))}{\lambda \prod_{j=1}^{\lambda_1} (\zeta_q(\theta) - \beta_j)^{m_j - 1}}.$$

Then

$$(10.5)\quad h_p(\theta', \theta) = \sum_{q=1}^{\lambda} f_q(\theta) \chi_{p,q}(\theta', \theta),$$

$$(10.6)\quad h_p(\theta', \theta) - f = \sum_{q=1}^{\lambda} (f_q(\theta) - f) \chi_{p,q}(\theta', \theta).$$

For a fixed value of θ', formulas (10.5) and (10.6) are true for almost every value of θ; and if it should happen for a given value of θ that they are true for some value of θ', then they are true for all other values of θ'.

The Jacobi series for $H(Z, \omega(z)/\mu)$ with respect to the points α_j is the series

$$(10.7)\quad \sum_{\nu=0}^{\infty} \{a_{\lambda\nu} + a_{\lambda\nu+1}(Z - \alpha_1) + \cdots + a_{\lambda\nu+\lambda-1}(Z - \alpha_1) \cdots (Z - \alpha_{\lambda-1})\} [\omega(z)]^{\nu}.$$

The Taylor series for $H(Z, w)$ about the point $w=0$ is obtained from (10.7) by substituting μw for $\omega(z)$.

THEOREM 10.3. *If $F(z) \in H_\Gamma$, the series*

$$\sum_{\nu=0}^{\infty} \{a_{\lambda\nu} + a_{\lambda\nu+1}(Z - \alpha_1) + \cdots + a_{\lambda\nu+\lambda-1}(Z - \alpha_1) \cdots (Z - \alpha_{\lambda-1})\} \mu^{\nu} e^{i\nu\theta}$$

is the Fourier series for the function $H(Z, e^{i\theta})$, considered as a function of θ.

The theorem is a consequence of (10.7), Theorem 10.1 and the theorem of F. and M. Riesz to which allusion has previously been made [15, p. 42].

Henceforth we assume invariably that $F(z) \in H_\Gamma$.

We write:

$$S_n(z_p(e^{i\theta})) = s_{p,n}(\theta),$$

$$\sum_{\nu=0}^{n} \{a_{\lambda\nu} + a_{\lambda\nu+1}(\zeta_p(\theta') - \alpha_1) + \cdots + a_{\lambda\nu+\lambda-1}(\zeta_p(\theta') - \alpha_1) \cdots (\zeta_p(\theta') - \alpha_{\lambda-1})\} \mu^{\nu} e^{i\nu\theta}$$

$$= \sum_{\nu=0}^{n} c_{p,\nu}(\theta') e^{i\nu\theta} = \sigma_{p,n}(\theta', \theta);$$

$$\sigma_{p,n}(\theta, \theta) = \sigma_{p,n}(\theta) = s_{p,\lambda n+\lambda-1}(\theta).$$

Dirichlet's integral formulas [22, pp. 20–21] for $\sigma_{p,n}(\theta', \theta)$ take the following forms:

$$(10.8)\qquad \sigma_{p,n}(\theta', \theta) = \frac{1}{\pi}\int_{-\pi}^{\pi} h_p(\theta', \theta + \tau)K_n(\tau)d\tau,$$

$$(10.9)\qquad \sigma_{p,n}(\theta', \theta) - f = \frac{1}{\pi}\int_{0}^{\pi} [h_p(\theta', \theta + \tau) + h_p(\theta', \theta - \tau) - 2f]K_n(\tau)d\tau,$$

both formulas holding for all θ and θ'. When $\theta'=\theta$, these formulas become useful expressions for the $\lambda(n+1)$th partial sum of the series (1.1). We obtain the following result by referring to Theorem 8.2:

THEOREM 10.4. *If* $\lim_{n\to\infty} \sigma_{p,n}(\theta)=f(\theta)$ *uniformly on a set E, then* $\lim_{n\to\infty} s_{p,n}(\theta)=f(\theta)$ *uniformly on E.*

11. **Convergence tests.** When $\bar{m}=1$, the integrals (10.8) and (10.9) are essentially no more complicated than the ordinary Dirichlet integral in the theory of Fourier series. However, when $\bar{m}>1$ the integrals must be studied by methods which are modifications or extensions of certain of the methods of Fourier analysis. Some of the extensions are trivial, some non-trivial and of interest in themselves.

The methods are perhaps best illustrated by deriving convergence criteria for the Jacobi series analogous to the de la Vallée Poussin test and to some of the tests which it includes. Accordingly, we shall devote this final section to such derivations.

It will be convenient to have formal statements of a few simple results on functions of bounded variation. If $\phi(t)$ is a complex-valued function of the real variable t which is of bounded variation on the interval $[0, t]$, we shall denote the absolute (or total) variation [16, p. 96](25) of $\phi(t)$ over the interval $[0, t]$ by $V(\phi; t)$. Obviously $V(\mathfrak{R}(\phi); t)\leq V(\phi; t)$ and $V(\mathfrak{J}(\phi); t)\leq V(\phi; t)$.

LEMMA 11.1. *Let* $f(t, \theta)$ *be a real function of the real variables t and* θ, *and let* $f(0, \theta)=0$. *A necessary and sufficient condition that* $f(t, \theta)$ *be of bounded variation in t for* $0\leq t\leq T$ *and for each* θ *on a set of E, and that* $\lim_{t\to 0} f(t, \theta)=\lim_{t\to 0} V(f; t)=0$ *uniformly for* $\theta\in E$, *is that* $f(t, \theta)$ *may be expressed as the difference of two non-negative functions of t and* θ *which are monotone non-decreasing functions of t for any fixed values of* $\theta\in E$, *and which approach zero uniformly,* $\theta\in E$, *as* $t\to 0$.

The sufficiency is obvious, and the necessity may be proved by simply writing down the usual expression for $f(t, \theta)$ in terms of its upper and lower variations with respect to t [16, p. 98; 18, p. 356].

LEMMA 11.2. *Let* $f(t, \theta)$ *and* $g(t, \theta)$ *be complex functions of the real variables t and* θ, *defined and integrable with respect to t on the closed interval* $[0, T]$ *for each* θ *on a set E. If* $g(t, \theta)$ *and the function*

(25) We shall use Saks' terminology in referring to functions of bounded variation.

$$F(t,\theta)=\begin{cases}\dfrac{1}{t}\displaystyle\int_0^t f(\tau,\theta)d\tau, & t\neq 0,\\ f(0,\theta), & t=0,\end{cases}$$

are of bounded variation in t over the interval $[0, T]$ for each $\theta\in E$, then so is the function

$$G_1(t,\theta)=\begin{cases}\dfrac{1}{t}\displaystyle\int_0^t g(\tau,\theta)f(\tau,\theta)d\tau, & t\neq 0,\\ g(+0,\theta)f(0,\theta), & t=0.\end{cases}$$

Moreover if

$$|F(t,\theta)|+V(F;t)\leqq M(F),\qquad |g(t,\theta)|+V(g;t)\leqq M(g),$$

for t on the interval $[0, T]$, $\theta\in E$, then

$$|G(t,\theta)|\leqq AM(F)M(g),\qquad V(G;t)\leqq AM(F)M(g),$$

for t on the interval $[0, T]$, $\theta\in E$, where A is an absolute constant. In particular, if $\lim_{t\to 0}F(t,\theta)=\lim_{t\to 0}V(F;t)=0$, or if $\lim_{t\to 0}g(t,\theta)=\lim_{t\to 0}V(g;t)=0$ uniformly, $\theta\in E$, then $\lim_{t\to 0}G(t,\theta)=\lim_{t\to 0}V(G;t)=0$ uniformly for $\theta\in E$.

It is sufficient to prove the theorem for the case in which $T>0$ and $F(t,\theta)$ and $g(t,\theta)$ are real, non-negative, non-decreasing functions of t for each $\theta\in E$. We write

$$G(t,\theta)=\frac{1}{t}\int_0^t gFd\tau+\frac{1}{t}\int_0^t g\frac{\partial F}{\partial\tau}\tau d\tau=I_1+I_2.$$

A well known theorem [17, p. 100] enables us to assert immediately I_1 is non-decreasing on the interval $[0, T]$; and clearly $I_1\leqq M(F)M(g)$. We integrate by parts in I_2:

$$I_2=\int_0^t g\frac{\partial F}{\partial\tau}d\tau-\frac{1}{t}\int_0^t\left[\int_0^u g\frac{\partial F}{\partial\tau}d\tau\right]du.$$

Each of these expressions is a non-decreasing function of t, and $\int_0^t g[\partial F/\partial\tau]d\tau \leqq M(g)[F(t,\theta)-F(0,\theta)]\leqq M(g)M(F)$. We leave the remaining details to the reader.

LEMMA 11.3. *Let $f(t,\theta)$ be defined as in Lemma 11.2. If $\lim_{t\to 0}\int_0^t|f(\tau,\theta)|d\tau=0$ uniformly, $\theta\in E$, then the function*

$$\phi(t,\theta)=\begin{cases}\dfrac{1}{t}\displaystyle\int_0^t \tau f(\tau,\theta)d\tau, & t\neq 0,\\ 0, & t=0,\end{cases}$$

is of bounded variation in t over the interval $[0, T]$ *for each* $\theta \in E$, *and*

$$\lim_{t\to 0} \phi(t, \theta) = \lim_{t\to 0} \frac{1}{t}\int_0^t \tau \,|\, f(\tau, \theta)\,|\, d\tau = \lim_{t\to 0} V(\phi; t) = 0$$

uniformly, $\theta \in E$.

A proof may be given by the methods used in the proof of the preceding lemma; we again leave the details to the reader.

We now let $f(\theta)$ be an arbitrary complex single-valued function of θ, and we write

$$H_p(\theta, \tau; f(\theta)) = h_p(\theta, \theta + \tau) + h_p(\theta, \theta - \tau) - 2f(\theta),$$

$$J_p(t, \theta; f(\theta)) = J_p(t) = \begin{cases} \dfrac{1}{t}\displaystyle\int_0^t H_p(\theta, \tau; f(\theta))d\tau, & t \neq 0, \\ 0, \quad t = 0. \end{cases}$$

This equation defines a single-valued function of t and θ for all values of θ for which $f(\theta)$ is defined, and for all values of t.

If a function $f(t, \theta)$, defined as in Lemma 11.2, is of bounded variation in t, $0 \leqq t \leqq T$, $T > 0$, for each θ on a set E, where T is independent of θ, and if $\lim_{t\to 0} f(t, \theta) = \lim_{t\to 0} V(f; t) = 0$ uniformly, $\theta \in E$, then we shall say that this function has *property $\mathcal{A}$ uniformly on E*. If E consists of a single point, we shall simply say that $f(t, \theta)$ has *property $\mathcal{A}$*[28].

The fundamental theorem of this section will now be stated:

THEOREM 11.1. (a) *If for some number f, $J_p(t, \theta; f)$ has property $\mathcal{A}$, then* $\lim_{n\to\infty} s_{p,n}(\theta) = f$.

(b) *If $f(\theta)$ is defined on a set E, if $|f(\theta)| \leqq M$, $\theta \in E$, and if $J_p(t, \theta; f(\theta))$ has property $\mathcal{A}$ uniformly on E, then* $\lim_{n\to\infty} s_{p,n}(\theta) = f(\theta)$ *uniformly on E.*

To prove the theorem, we first establish a result of the Riemann-Lebesgue type.

LEMMA 11.4. *Assume that $\phi(t)$ is a bounded integrable function of t for $a \leqq t \leqq b$. Then*

$$\lim_{n\to\infty} \int_a^b h_p(\theta', \theta + \tau)\phi(\tau) \begin{Bmatrix} \sin n\tau \\ \cos n\tau \end{Bmatrix} d\tau = 0, \qquad p = 1, \cdots, \lambda,$$

uniformly for all values of θ and θ'.

For we have

(28) It is perhaps worth while to point out that if $f(t, \theta)$ is of bounded variation in t, $0 \leqq t \leqq T$, $T > 0$, and if $\lim_{t\to 0} f(t, \theta) = 0$, then $\lim_{t\to 0} V(f; t) = 0$; but uniformity of the first limit in θ does not imply uniformity of the second limit in θ.

$$\int_a^b h_p(\theta', \theta + \tau)\phi(\tau)\begin{Bmatrix}\sin n\tau\\ \cos n\tau\end{Bmatrix}d\tau$$

$$= \int_a^b \phi_0(\theta + \tau)\phi(\tau)\begin{Bmatrix}\sin n\tau\\ \cos n\tau\end{Bmatrix}d\tau + (\zeta_p(\theta') - \alpha_1)\int_a^b \phi_1(\theta + \tau)\phi(\tau)\begin{Bmatrix}\sin n\tau\\ \cos n\tau\end{Bmatrix}d\tau$$

$$+ \cdots + (\zeta_p(\theta') - \alpha_1)\cdots(\zeta_p(\theta') - \alpha_{\lambda-1})\int_a^b \phi_{\lambda-1}(\theta + \tau)\phi(\tau)\begin{Bmatrix}\sin n\tau\\ \cos n\tau\end{Bmatrix}d\tau,$$

and we may apply a familiar extension of the Riemann-Lebesgue theorem [22, p. 22] to each of the terms in this sum, which are periodic functions of θ and θ'.

Thus for any fixed δ, $0<\delta<\pi$, we find that

$$\lim_{n\to\infty}\int_\delta^\pi H_p(\theta, \tau; f)K_n(\tau)d\tau = 0 \tag{11.1}$$

uniformly, $\theta\in E$. We write

$$\int_0^\delta H_p(\theta, \tau; f)K_n(\tau)d\tau = \int_0^\delta J_p(\tau)K_n(\tau)d\tau + \int_0^\delta J_p'(\tau)\tau K_n(\tau)d\tau.$$

We apply Lemma 11.1 to the real and imaginary parts of $J_p(t)$; and by proceeding exactly as in the classical proof of Dirichlet's theorem [22, pp. 25–26; 18, p. 407], we can show that for an arbitrary positive number ϵ, there exists a number δ' independent of n and θ such that if $\delta<\delta'$

$$\left|\int_0^\delta J_p(\tau)K_n(\tau)d\tau\right| < \epsilon, \quad \theta\in E, \ n = 0, 1, 2, \cdots. \tag{11.2}$$

Suppose now that $j(t)$ is one of the monotone components of $J_p(t)$ referred to in Lemma 11.1. Then there exists a number δ'' independent of n and θ such that if $\delta<\delta''$ then

$$\int_0^\delta j'(\tau)\tau K_n(\tau)d\tau \leqq C\int_0^\delta j'(\tau)d\tau \leqq Cj(\delta) < \epsilon, \ \theta\in E, n = 0, 1, \cdots, \tag{11.3}$$

where C is an absolute constant. Inequalities similar to (11.3) may be established for each of the other monotone components of $J_p(t)$. The proof is completed by combining these inequalities with (11.1) and (11.2), and by applying Theorem 10.4.

The theorems in the remainder of this section contain sufficient conditions for convergence which are imposed directly upon the function $F(z)$ or upon its transform on the surface R.

The first of these theorems is an exact analogue of de la Vallée Poussin's test.

THEOREM 11.2. *If $\zeta_{p_1}(\xi)=\zeta_{p_2}(\xi)=\cdots=\zeta_{p_m}(\xi)$, $m\geqq 1$, and if numbers f_h, $h=1,\cdots,m$, can be chosen so that the functions*

$$K_{p_h}(t)=\begin{cases}\dfrac{1}{t}\displaystyle\int_0^t (f_{p_h}(\xi+\tau)+f_{p_h}(\xi-\tau)-2f_h)d\tau, \ t\neq 0,\\ 0, \ t=0,\end{cases} \qquad h=1,\cdots,m,$$

all have property $\mathcal{A}$, then $\lim_{n\to\infty} s_{p_1,n}(\xi)=\bar{f}$, where $\bar{f}=\sum_1^m f_h/m$.

We give the proof by establishing the following lemma:

LEMMA 11.5. *If $K_{p_h}(t)$ has property $\mathcal{A}$, $h=1,\cdots,m$, then $J_{p_1}(t,\xi;\bar{f})$ has the property $\mathcal{A}$.*

We write

$$\begin{aligned}
J_{p_1}(t,\xi;\bar{f}) &= \sum_{h=1}^m \frac{1}{t}\int_0^t [f_{p_h}(\xi+\tau)+f_{p_h}(\xi-\tau)-2f_h]\chi_{p_1,p_h}(\xi,\xi+\tau)d\tau\\
&\quad+\sum_{h=1}^m\frac{1}{t}\int_0^t\frac{f_{p_h}(\xi-\tau)-f_h}{\omega_{p_h}(\xi-\tau)}\left[\frac{\chi_{p_1,p_h}(\xi,\xi-\tau)-\chi_{p_1,p_h}(\xi,\xi+\tau)}{\tau^{1/m}}\right]\\
&\qquad\cdot\frac{\omega_{p_h}(\xi-\tau)}{\tau^{(m-1)/m}}\tau d\tau\\
&\quad+\sum_{h=1}^m\frac{1}{t}\int_0^t(f_h-\bar{f})[\chi_{p_1,p_h}(\xi,\xi+\tau)+\chi_{p,p_h}(\xi,\xi-\tau)]d\tau\\
&\quad+\sum{}'\frac{1}{t}\int_0^t[(f_q(\xi+\tau)-\bar{f})\chi_{p_1,q}(\xi,\xi+\tau)\\
&\qquad\qquad+(f_q(\xi-\tau)-\bar{f})\chi_{p_1,q}(\xi,\xi-\tau)]d\tau\\
&=\sum_{h=1}^m I_h^{(1)}+\sum_{h=1}^m I_h^{(2)}+\sum_{h=1}^m I_h^{(3)}+\sum{}' I_q^{(4)}, \qquad t\neq 0,
\end{aligned}\tag{11.4}$$

where $\sum'=\sum_{q=1}^{\lambda}$, $q\neq p_1, p_2,\cdots,p_m$[27]. If $m>1$, then for some subscript j, we have $\xi\equiv\xi_j \pmod{2\pi}$, $\zeta_{p_1}(\xi)=\beta_j$, and $m=m_j$; and

$$\chi_{p_1,q}(\xi,\xi+\tau)=\frac{\prod'[\zeta_q(\xi+\tau)-\zeta_n(\xi)]}{\lambda\prod''[\zeta_q(\xi+\tau)-\beta_n]^{m_n-1}},$$

where $\prod'=\prod_{n=1}^{\lambda}$, $n\neq p_1,p_2,\cdots,p_m$, and $\prod''=\prod_{n=1}^{\lambda_1}$, $n\neq j$. Thus the function $\chi_{p_1,p_h}(\xi,\xi+\tau)$ is a rational function of $\zeta_{p_h}(\xi+\tau)$ with no poles for $-\delta_j\leqq\tau\leqq\delta_j$, and consequently is an absolutely continuous function of τ in this interval. Applying Lemma 11.2, we find that $I_h^{(1)}$ is of bounded variation in an interval to the right of $t=0$, and $\lim_{t\to 0} I_h^{(1)}=0$.

Now

[27] We may write $I_p^{(k)}=0$ for $t=0$.

$$|\chi_{p_1,p_h}(\xi,\xi-\tau)-\chi_{p_1,p_h}(\xi,\xi+\tau)|\leqq M|\zeta_{p_h}(\xi-\tau)-\zeta_{p_h}(\xi+\tau)|\leqq M_1|\tau|^{1/m}$$

for $|\tau|\leqq\delta_j$, according to (6.7); and, as we have seen in the proof of Lemma 6.1, $|\omega_{p_h}(\xi-\tau)|/|\tau|^{(m-2)/m}$ is uniformly bounded for $|\tau|\leqq\delta_j$. Therefore, since $\lim_{t\to 0}\int_0^t|(f_{p_h}(\xi-\tau)-f_h)/\omega_{p_h}(\xi-\tau)|d\tau=0$, we may apply Lemma 11.3 to $I_h^{(2)}$, and we find that it is of bounded variation for $|t|\leqq\delta_j$ and $\lim_{t\to 0}I_h^{(2)}=0$.

Next we notice that

$$\lim_{t\to 0}\chi_{p_1,p_h}(\xi,\xi+t)=\lim_{z\to\beta_j}\frac{\omega(z)-\omega(\beta_j)}{\omega'(z)(z-\beta_j)}=\frac{1}{m},\qquad h=1,\cdots,m,$$

where $z=\zeta_{p_h}(\xi+t)$. Since $\sum_{h=1}^m(f_h-\bar f)=0$, we find from Lemma 11.2 that $I_h^{(3)}$ is of bounded variation for $|t|\leqq\delta_j$ and $\lim_{t\to 0}I_h^{(3)}=0$.

In connection with $I_q^{(4)}$, we shall state for future reference a result which is rather stronger than the one needed at present[28].

LEMMA 11.6. *If $\phi(\theta)$ is a function such that $\phi(\theta)/\omega_q(\theta)$ is integrable on an interval $[a, b]$, $a<b$, and if a number T exists, $0<T\leqq(b-a)/2$, such that $|\zeta_p(\theta)-\zeta_q(\theta+t)|$ is bounded from zero for all θ on the interval $[a+T, b-T]$, and for $-T\leqq t\leqq T$, then the function*

$$\Phi_{p,q}(t)=\Phi_{p,q}(t,\theta;\phi)=\begin{cases}\dfrac{1}{t}\displaystyle\int_0^t\phi(\theta+\tau)\chi_{p,q}(\theta,\theta+\tau)d\tau, & t\neq 0,\\ 0, \quad t=0,\end{cases}$$

is of bounded variation in t for $-T\leqq t\leqq T$ and for each θ in the interval $[a+T, b-T]$, and

$$\begin{aligned}\lim_{t\to 0}\Phi_{p,q}(t)&=\lim_{t\to 0}\frac{1}{t}\int_0^t|\phi(\theta+\tau)\chi_{p,q}(\theta,\theta+\tau)|d\tau\\&=\lim_{t\to 0}V(\Phi_{p,q};t)=0\end{aligned}$$

uniformly for θ on the interval $[a+T, b-T]$.

The proof of this lemma may be given by referring to Lemma 11.3, in which we let

$$f(t,\theta)=\frac{\phi(\theta+t)}{\omega_q(\theta+t)}\cdot\frac{1}{\zeta_q(\theta)-\zeta_q(\theta+t)}\cdot\frac{\mu e^{i\theta}-\mu e^{i(\theta+t)}}{t}.$$

The application of Lemma 11.6 to $I_q^{(4)}$ is obvious. The proof of Lemma 11.5

(28) Lemmas 11.6 and 11.8 are stated in a form in which they are useful in obtaining certain results concerning the Cesàro means of the Jacobi series which the author intends to publish elsewhere. Thus the functions $Y_{p,q}(t,\theta)$ and $y(t,\theta)$ in Lemma 11.8 and its proof are not needed in the present discussion.

for $m>1$ is completed by the observation that the sum of a finite number of functions which have property $\mathcal{A}$, itself has property $\mathcal{A}$; we have shown $J_{p_1}(t, \xi; \bar{f})$ to be such a sum.

The case $m=1$ is treated similarly, but the details are simpler, and will be omitted.

Using the notation of this theorem, we may state a test analogous to Dini's test:

THEOREM 11.3. *If the functions*

$$\left| \frac{f_{p_h}(\xi + t) + f_{p_h}(\xi - t) - 2f_h}{t} \right|, \qquad h = 1, \cdots, m,$$

are integrable in an interval to the right of $t=0$, *then* $\lim_{n\to\infty} s_{p_1,n}(\xi) = \bar{f}$.

A proof may be given by showing that the conditions in the hypothesis of Theorem 11.2 are satisfied; this may be conveniently done by using the method usually employed to show that de la Vallée Poussin's test includes Dini's test [18, p. 410; 22, p. 36]. The theorem may also be proved by breaking up the integral in (10.9) into a number of integrals which correspond to the functions $I_h^{(1)}, I_h^{(2)}, I_h^{(3)}, I_q^{(4)}$ of the proof of Theorem 11.2, and by showing that the Riemann-Lebesgue theorem applies to each one.

The following result can easily be established by using Theorem 11.3.

THEOREM 11.4. *Let* β *be a point of* Γ *of multiplicity* $m, m \geqq 1$; *belonging to the* m *curves* $C_{k_1}, C_{k_2}, \cdots, C_{k_m}$. *If* $|F(z)-F_h| \leqq M|z-\beta|^\eta$, $\eta>0$, *for almost every* $z \in C_{k_h}$ *in some neighborhood of* β, $h=1, \cdots, m$, *then* $\lim_{n\to\infty} S_n(\beta) = \sum_1^m F_h/m$.

Another consequence of Theorem 11.2 worth mentioning is the following theorem.

THEOREM 11.5. *Let* β *be a point of* Γ *of multiplicity* m, $m \geqq 1$, *belonging to the* m *curves* $C_{k_1}, C_{k_2}, \cdots, C_{k_m}$. *If* $F(z)$ *is of bounded variation on some arc of* C_{k_h} *containing* β *as an interior point and if* $\lim_{z\to\beta} F(z) = F_h$, $z \in C_{k_h}$, $h=1, \cdots, m$, *then* $\lim_{n\to\infty} S_n(\beta) = \sum_1^m F_h/m$.

The proof may be given by using Lemma 11.2 to show that the conditions in the hypothesis of Theorem 11.2 are satisfied by the numbers F_h and by the transforms on R of the function $F(z)$.

We turn now to the question of uniform convergence. By an arc of Γ, we shall mean a closed arc. A proper subarc of an arc A will be a closed subarc which contains neither end-point of A.

Our first theorem is the analogue of Jordan's test.

THEOREM 11.6. *Let* β *be a point of* Γ *of multiplicity* m, $m \geqq 1$, *belonging to the* m *curves* $C_{k_1}, C_{k_2}, \cdots, C_{k_m}$. *Let* A_h *be an arc of* C_{k_h} *containing* β *as an interior point, but containing no point of* Γ *of multiplicity greater than one, other*

than β. If $F(z)$ is single-valued, continuous[29], *and of bounded variation on the set $\sum_{h=1}^{m} 1A_h$, then $\lim_{n\to\infty} S_n(z)=F(z)$ uniformly on any proper subarc of A_h, $h=1, \cdots, m$. If for any given h, $A_h=C_{k_h}$, then $\lim_{n\to\infty} S_n(z)=F(z)$ uniformly on C_{k_h}.*

We base the proof on Theorem 11.1(b) and on certain properties of the functions $\chi_{p,q}(\theta, \theta+t)$, which we study by means of two lemmas. For future reference, we state these results in a somewhat more precise form than is necessary for present purposes.

Let

$$S_\alpha(t, \theta) = S(t, \theta) = \begin{cases} \dfrac{1}{t}\displaystyle\int_0^t \left| \dfrac{\sin \frac{1}{2}\theta}{\sin \frac{1}{2}(\theta+\tau)} \right|^\alpha d\tau, & 0 < \alpha < 1,\ t \neq 0, \\ 1, \quad t = 0, \end{cases}$$

$$\sigma_\alpha(t, \theta) = \sigma(t, \theta) = \begin{cases} \dfrac{1}{t}\displaystyle\int_0^t \left| \dfrac{\theta}{\theta+\tau} \right|^\alpha d\tau, & 0 < \alpha < 1,\ t \neq 0, \\ 1, \quad t = 0. \end{cases}$$

LEMMA 11.7. *The functions $S(t, \theta)$, $V(S; t)$, $\sigma(t, \theta)$, and $V(\sigma; t)$ are uniformly bounded, $|\theta| \leqq \frac{1}{2}\pi$, $|t| \leqq \frac{1}{2}\pi$.*

For the proof, we observe first that the function

$$\left| \frac{\sin \frac{1}{2}(\theta)}{\theta} \cdot \frac{\theta+t}{\sin \frac{1}{2}(\theta+t)} \right|^\alpha$$

and its abolute variation with respect to t are uniformly bounded for $|\theta| \leqq \frac{1}{2}\pi$, $|t| \leqq \frac{1}{2}\pi$, so by Lemma 11.2, it will suffice to establish the desired result for the function $\sigma(t, \theta)$. The result is trivial for $\theta=0$; henceforth we suppose that $\theta\neq 0$.

$$\sigma(t, \theta) = \frac{\theta}{t}\int_0^{t/\theta} \frac{1}{|1+u|^\alpha}\, du = \frac{1}{U}\int_0^{U} \frac{1}{|1+u|^\alpha}\, du = f(U),$$

where $U=t/\theta$. The function $f(U)$ is continuous for all values of U, and is a monotone increasing function of U for $U<-1$ and a monotone decreasing function of U for $U>-1$. Thus for any fixed value of θ, $V(S; t)$ is equal either to $|\sigma(t, \theta)-\sigma(0, \theta)|$ or to $\sigma(-\theta, \theta)-\sigma(0, \theta)+\sigma(-\theta, \theta)-\sigma(t, \theta)$, and since $\sigma(t, \theta)$ is obviously uniformly bounded for all values of t and θ, so also is $V(S; t)$. The proof is complete.

Now let

(29) If $F(z)$ is of bounded variation on an arc of Γ, it can have only simple discontinuities on that arc. It can be shown that such discontinuities are impossible on each of the curves C_k; therefore this condition of continuity may be replaced by the requirement that $F(z)$ be single-valued for $z=\beta$.

$$X_{p,q}(t,\theta)=\begin{cases}\dfrac{1}{t}\displaystyle\int_0^t \chi_{p,q}(\theta,\theta+\tau)d\tau, & t\neq 0,\\ 0, & t=0,\end{cases}\qquad p=1,\cdots,\lambda,\ q=1,\cdots,\lambda,$$

and

$$Y_{p,q}(t,\theta)=\begin{cases}\dfrac{1}{t}\displaystyle\int_0^t |\chi_{p,q}(\theta,\theta+\tau)|d\tau, & t\neq 0,\\ 0, & t=0,\end{cases}\qquad p=1,\cdots,\lambda,\ q=1,\cdots,\lambda.$$

LEMMA 11.8. *If T is any positive number, there exists a number $M(T)$ independent of θ and t such that $|X_{p,q}(t,\theta)|\leq Y_{p,q}(t,\theta)\leq M(T)$, $V(X_{p,q};t)\leq M(T)$, $p=1,\cdots,\lambda$, $q=1,\cdots,\lambda$, for $|t|\leq T$ and for all values of θ.*

The functions $X_{p,q}$ and $Y_{p,q}$ are each periodic functions of θ, so for a given p and q it will suffice to prove the lemma for the values of θ in a period. Now the numerator of the third member of (10.4) is a polynomial in $\zeta_q(\theta)$ and $\zeta_j(\theta')$, $j=1,\cdots,\lambda$, $j\neq p$, so this numerator is bounded in modulus for all θ and θ'. Therefore, since $\int_0^t|\omega_q(\theta+\tau)|^{-1}d\tau$ is uniformly bounded for all θ and for $|t|\leq T$, so also is $\int_0^t|\chi_{p,q}(\theta,\theta+\tau)|d\tau$. Furthermore,

$$V(t\cdot X_{p,q};t)\leq\int_{-T}^{T}|\chi_{p,q}(\theta,\theta+\tau)|d\tau$$

for $|t|\leq T$ and for all θ. It follows from these remarks that given any δ, $0<\delta<T$, a positive number $M(\delta)$ independent of θ exists such that

$$|X_{p,q}(t,\theta)|\leq Y_{p,q}(t,\theta)\leq M(\delta),\quad V(X_{p,q};t)\leq M(\delta),\quad \delta\leq|t|\leq T,\quad \text{all }\theta.$$

We now prove that for each real number ξ, there exist positive numbers $M(\xi)$ and $\delta(\xi)$ independent of t and θ such that $Y_{p,q}(t,\theta)\leq M(\xi)$ and $V(X_{p,q};t)\leq M(\xi)$ for $|t|\leq\delta(\xi)$ and for θ in the interval $I(\xi)$: $\xi-\delta(\xi)\leq\theta\leq\xi+\delta(\xi)$. If $\omega_q'(\xi)\neq 0$, then it may easily be shown (compare the proof of Theorem 11.2) that both $|\chi_{p,q}(\theta,\theta+t)|$ and $V[\chi_{p,q};t]$ are uniformly bounded for t suitably restricted and for θ in some interval containing ξ in its interior. The existence of $\delta(\xi)$ and $M(\xi)$ then follows from Lemma 11.2.

Suppose instead that $\xi\equiv\xi_i \pmod{2\pi}$ and $\zeta_{q_1}(\xi)=\zeta_{q_2}(\xi)=\cdots=\zeta_{q_m}(\xi)=\beta$. Let $\delta(\xi)=\delta_i/2$. If $p\neq q_h$, $h=1,\cdots,m$, the existence of $M(\xi)$ is again immediate, this time because of Lemma 11.6 (with $\phi(\theta)=1$). If $p=q_H$, $1\leq H\leq m$, and $q=q_K$, $1\leq K\leq m$, we write

$$\begin{aligned}\chi_{p,q}(\theta,\theta+t)&=\frac{\prod_h'(\zeta_q(\theta+t)-\zeta_{q_h}(\theta))}{(\zeta_q(\theta+t)-\beta)^{m-1}}R(\theta,\theta+t)\\&=R(\theta,\theta+t)\prod_h{}'\left[1-\frac{\zeta_{q_h}(\theta)-\beta}{\zeta_q(\theta+t)-\beta}\right]\end{aligned}$$

where $\prod_h' = \prod_{h=1}^m$, $h \neq H$, and where $R(\theta, \theta+t)$ is a rational function of $\zeta_q(\theta+t)$ with constant coefficients in the denominator, and with numerator coefficients which are polynomials in certain of the functions $\zeta_p(\theta)$. Moreover $R(\theta, \theta+t)$, considered as a function of $\zeta_q(\theta+t)$, has no poles for $|\theta-\xi| \leq \frac{1}{2}\delta_j$ and $|t| \leq \frac{1}{2}\delta_j$. Consequently, $|R(\theta, \theta+t)|$ and $V(R; t)$ are uniformly bounded for θ in $I(\xi)$ and $|t| \leq \delta(\xi)$. By reference to Lemma 11.2, it is now seen that our problem reduces to one of showing that if we let

$$x(t, \theta) = \begin{cases} \dfrac{1}{t}\displaystyle\int_0^t \prod_h{}'' \left[\frac{\zeta_{q_h}(\theta) - \beta}{\zeta_q(\theta + \tau) - \beta}\right] d\tau, & t \neq 0, \\ 1, \quad t = 0, \end{cases}$$

$$y(t, \theta) = \begin{cases} \dfrac{1}{t}\displaystyle\int_0^t \prod_h{}'' \left|\frac{\zeta_{q_h}(\theta) - \beta}{\zeta_q(\theta + \tau) - \beta}\right| d\tau, & t \neq 0, \\ 1, \quad t = 0, \end{cases}$$

where in $\prod_h''$, the index h assumes σ values which are an arbitrary subset (not containing the integer H) of the first m positive integers—then $y(t, \theta)$ and $V(x; t)$ are uniformly bounded for θ in $I(\xi)$ and for $|t| \leq \delta(\xi)$.

We use (6.5) to write

$$\begin{aligned} x(t, \theta) &= \frac{1}{t}\int_0^t \left|\frac{\sin \frac{1}{2}(\theta - \xi)}{\sin \frac{1}{2}(\theta - \xi + \tau)}\right|^{\sigma/m} \Bigg[e^{(i\sigma/m)(\psi(\theta)-\psi(\theta+\tau))} \\ &\qquad \cdot e^{2\pi i\Delta/m}\frac{\prod_h'' A(W_h)}{[A(\overline{W})]^\sigma}\Bigg] d\tau, \qquad t \neq 0, \\ y(t, \theta) &= \frac{1}{t}\int_0^t \left|\frac{\sin \frac{1}{2}(\theta - \xi)}{\sin \frac{1}{2}(\theta - \xi + \tau)}\right|^{\sigma/m} \frac{|\prod_h'' A(W_h)|}{|A(\overline{W})|^\sigma} d\tau, \qquad t \neq 0, \end{aligned} \tag{11.5}$$

where $\overline{W} = (e^{i(\theta+\tau)} - e^{i\xi})_K^{1/m}$, and Δ is a constant independent of θ and t. The quantity in brackets in (11.5) is uniformly bounded in modulus for θ in $I(\xi)$ and for $|\tau| \leq \delta(\xi)$, and so also is its absolute variation with respect to τ (compare proof of Theorem 8.4). Lemmas 11.2 and 11.7 therefore permit us to draw the desired conclusions concerning $x(t, \theta)$ and $y(t, \theta)$.

The proof of the lemma may now be completed by an appeal to the Heine-Borel theorem.

We now proceed with the proof of Theorem 11.6. Let us suppose (as obviously we may) that the functions $z_p(w)$ are numbered in such a way that $z = \zeta_h(\theta)$, $\theta_h \leq \theta \leq \theta_h'$, $h = 1, \cdots, m$, are the equations respectively of the arcs A_h, $h = 1, \cdots, m$. We choose the intervals I_h: $[\theta_h, \theta_h']$ so that each contains the number ξ such that $\zeta_1(\xi) = \zeta_2(\xi) = \cdots = \zeta_m(\xi) = \beta$. Since the numbering of the curves C_{k_h} is quite arbitrary, it will evidently suffice to prove that $\lim_{n\to\infty} s_{1,n}(\theta) = f_1(\theta)$ uniformly on the closed interval $I(\delta)$: $[\theta_1 + \delta, \theta_1' - \delta]$, where δ is any positive number less than the smaller of the two numbers

$(\theta_1'-\xi)/2$ and $(\xi-\theta_1)/2$. This will be done by proving the following result.

LEMMA 11.9. *The hypotheses of Theorem* 11.6 *imply that* $J_1(t,\theta;f_1(\theta))$ *has property* $\mathcal{A}$ *uniformly on* $I(\delta)$.

As we remarked in the proof of Theorem 8.4, $f_h(\theta)$ is continuous, and of bounded variation on θ for $\theta_h\leqq\theta\leqq\theta_h'$. We now write $J_1(t,\theta;f_1(\theta))=\sum_{h=1}^{\lambda}(P_h+Q_h)$ where

$$P_h=\begin{cases}\dfrac{1}{t}\displaystyle\int_0^t(f_h(\theta+\tau)-f_1(\theta))\chi_{1,h}(\theta,\theta+\tau)d\tau, & t\neq 0,\\ 0, \quad t=0;\end{cases}$$

$$Q_h=\begin{cases}\dfrac{1}{t}\displaystyle\int_0^t(f_h(\theta-\tau)-f_1(\theta))\chi_{1,h}(\theta,\theta-\tau)d\tau, & t\neq 0,\\ 0, \quad t=0.\end{cases}$$

It follows immediately from Lemmas 11.8 and 11.2 (in which we let $g(t,\theta)=f_1(\theta+t)-f_1(\theta))$ that P_1 and Q_1 are of bounded variation in t for θ in $I(\delta)$, and for $0\leqq t\leqq\delta$, and that $\lim_{t\to 0}P_1=\lim_{t\to 0}Q_1=\lim_{t\to 0}V(P_1;t)=\lim_{t\to 0}V(Q_1;t)=0$ uniformly for θ on $I(\delta)$.

Consider now P_H, $2\leqq H\leqq m$. Given any $\epsilon>0$, there exists a positive number $\delta(\epsilon)<\delta$ independent of θ and t such that if we write $g(t,\theta)=f_H(\theta+t)-f_1(\theta)=[f_H(\theta+t)-f_H(\xi)]+[f_H(\xi)-f_1(\theta)]$, then $|g(t,\theta)|<\epsilon$ and $V(g;t)<\epsilon$ for $|\theta-\xi|\leqq\delta(\epsilon)$, $|t|\leqq\delta(\epsilon)$. We find from Lemma 11.8 (in which we let $T=\delta$) and Lemma 11.2, that $|P_H|\leqq 4A\epsilon M(\delta)$ and $V(P_H;t)<4A\epsilon M(\delta)$ for $(\theta-\xi)\leqq\delta(\epsilon)$, $|t|\leqq\delta(\epsilon)$. On the other hand, if $|t|\leqq\delta(\epsilon)/2$ and if $\theta_1\leqq\theta\leqq\xi-\delta(\epsilon)$, $\xi+\delta(\epsilon)\leqq\theta\leqq\theta_1'$, then the quantity $|\zeta_H(\theta+t)-\zeta_1(\theta)|$ is bounded from zero, and also $|f_1(\theta)|$ is bounded. We then find from Lemma 11.6 (in which we let $T=\delta(\epsilon)/2$) that P_H is of bounded variation in t for these values of t and θ, and that furthermore, there exists a positive number $\delta'(\epsilon)\leqq\delta(\epsilon)/2$ such that $|P_H|<\epsilon$ and $V(P_H;t)<\epsilon$ for $|t|\leqq\delta'(\epsilon)$ and for $\theta_1\leqq\theta\leqq\xi-\delta(\epsilon)$, $\xi+\delta(\epsilon)\leqq\theta\leqq\theta_1'$. We have proved that P_H has property $\mathcal{A}$ uniformly on $I(\delta)$. Moreover, the argument has been stated so as to show simultaneously that Q_H has property $\mathcal{A}$ uniformly on $I(\delta)$.

Now $|\zeta_1(\theta)-\zeta_h(\theta+t)|\geqq\big||\zeta_h(\theta+t)-\zeta_1(\theta+t)|-|\zeta_1(\theta+t)-\zeta_1(\theta)|\big|$. If $h>m$, then there exist positive numbers d and $\delta''\leqq\delta$ independent of θ and t such that $|\zeta_h(\theta+t)-\zeta_1(\theta+t)|>d$ and $|\zeta_1(\theta+t)-\zeta_1(\theta)|<d/2$ for θ on $I(\delta)$ and $|t|\leqq\delta''$. Therefore, although the point $\zeta_h(\theta)$ may lie on the curve C_{k_1}, nevertheless $|\zeta_1(\theta)-\zeta_h(\theta+t)|$ is uniformly bounded from zero for all such values of θ and t. It follows from Lemma 11.6 that P_h and Q_h, $h>m$, have property $\mathcal{A}$ uniformly on $I(\delta)$.

We have now shown that $J_1(t,\theta;f_1(\theta))$ is the sum of a finite number of functions, each of which has property $\mathcal{A}$ uniformly on $I(\delta)$. It follows at once

that J_1 itself has property $\mathcal{A}$ uniformly on $I(\delta)$, and the proof of Lemma 11.9 is complete.

It follows immediately from Theorem 11.1(b) that $\lim_{n\to\infty} S_n(z)=F(z)$ uniformly on the subarc of A_1 which is the transform of $I(\delta)$.

If $A_1=C_{k_1}$, we may replace the interval $I(\delta)$ by the interval $[\theta_1, \theta_1']$ in the above argument, where now $\theta_1'-\theta_1=2\pi\kappa_{k_1}$. We then find that $\lim_{n\to\infty} S_n(z) =F(z)$ uniformly on C_{k_1}. The proof of the theorem is complete.

It is of interest to observe that if $\lim_{n\to\infty} S_n(z)=F(z)$ uniformly on C_k, then the sequence $\{S_n(z)\}$ converges uniformly in the closed limited Jordan region bounded by C_k, by the principle of the maximum. It follows in this case that $F(z)$ is uniformly continuous in this closed region[30].

We append two corollaries of Theorem 11.6; the second of these will be stated in terms of the notation introduced in the proof of that theorem.

THEOREM 11.7. *If $F(z)$ is continuous*[31] *and of bounded variation on Γ, then* $\lim_{n\to\infty} S_n(z)=F(z)$ *uniformly on $\overline{D}$.*

THEOREM 11.8. *If* (i) $m>1$, (ii) *$F(z)$ is single valued, continuous*[31], *and of bounded variation on A_1,* (iii) *P_h and Q_h have property $\mathcal{A}$ uniformly on I_h, $h=2, \cdots, m$, then* $\lim_{n\to\infty} S_n(z)=F(z)$ *uniformly on any proper subarc of A_1.*

The proof of Theorem 11.6 may also be given by a direct consideration of the integral in (10.9), but further lemmas of the Riemann-Lebesgue type are then needed to take care of the terms involving $f_h(\theta+\tau)$, $h>1$. Our development has indicated that the tests in Theorem 11.1(a) and (b) include all the other tests of this section. We add the remark that the tests contained in Theorems 11.3 and 11.5 are not comparable.

REFERENCES

1. G. A. Bliss, *Algebraic Functions*, New York, 1933.
2. E. Borel, *Leçons sur les Fonctions des Variables Réelles*, Paris, 1905.
3. T. J. I'A. Bromwich, *An Introduction to the Theory of Infinite Series*, 2d edition, London, 1926.
4. J. H. Curtiss, *A note on the Cesàro method of summation*, Bulletin of the American Mathematical Society, vol. 43 (1937), pp. 703–708.
5. M. Fekete, *Über den Schottkyschen Satz*, Journal für die reine und angewandte Mathematik, vol. 165 (1931), pp. 217–224.
6. G. Frobenius, *Über die Entwicklung analytischer Funktionen in Reihen, die nach gegebenen Funktionen fortschreiten*, ibid., vol. 73 (1871), pp. 1–30.
7. Édouard Goursat, *Cours d'Analyse*, vol. 2, part 1; English translation by E. R. Hedrick and Otto Dunkel, Boston, 1916.

(30) With reference to this remark, it is perhaps worth while to point out that the existence of a continuous boundary value function in the sense of §7 does not by itself imply that $F(z)$ is uniformly continuous in D. To draw this inference, additional restrictions are necessary, such as that $F(z) \in H_\Gamma$.

(31) See the footnote of Theorem 11.6.

8. C. G. J. Jacobi, *Über Reihenentwicklungen, welche nach den Potenzen eines gegebenen Polynoms fortschreiten, und zu Coefficienten Polynome eines niedereren Grades haben*, Journal für die reine und angewandte Mathematik, vol. 53 (1856–1857), pp. 103–126.

9. P. W. Ketchum, *On the expansion of a function analytic at distinct points*, Bulletin of the American Mathematical Society, vol. 43 (1937), pp. 115–121.

10. A. Kienast, *Über die Darstellung der analytischen Funktionen durch Reihen, die nach Potenzen eines Polynoms fortschreiten und Polynome eines niedereren Grades zu Koeffizienten haben*, Dissertation, Zurich, 1906.

11. P. Martinotti, *Su le serie d'interpolazione*, Rendiconti del'Istituto Lombardo, (2), vol. 43 (1910), pp. 391–401.

12. P. Montel, *Séries de Polynomes*, Paris, 1910, pp. 47–49, 95–97.

13. K. Rieder, *Polynomische Entwicklungen von Funktionen einer komplexen Variabeln*, Dissertation, Basel, 1911, pp. 66–85.

14. F. Riesz, *Über die Randwerte einer analytischen Funktion*, Mathematische Zeitschrift, vol. 18 (1923), pp. 87–95.

15. F. and M. Riesz, *Über Randwerte einer analytischen Funktion*, Quatrième Congrès des Mathematiciens Scandinaves, 1916, pp. 27–44.

16. S. Saks, *Theory of the Integral*, 2d revised edition, Warsaw, 1937.

17. J. D. Tamarkin, *The Theory of Fourier Series*, Providence, 1933 (mimeographed notes).

18. E. C. Titchmarsh, *The Theory of Functions*, Oxford, 1932.

19. J. L. Walsh, *Approximation by Polynomials in the Complex Domain*, Paris, 1935, pp. 13–14, 45–46.

20. J. L. Walsh, *Interpolation and Approximation by Rational Functions in the Complex Domain*, American Mathematical Society Colloquium Publications, vol. 20, New York, 1935.

21. J. L. Walsh, *Lemniscates and equipotential curves of Green's function*, American Mathematical Monthly, vol. 52 (1935), pp. 1–17.

22. A. Zygmund, *Trigonometrical Series*, Warsaw, 1935.

CORNELL UNIVERSITY,
ITHACA, N. Y.

ON SETS OF MATRICES WITH COEFFICIENTS IN A DIVISION RING

BY

RICHARD BRAUER

A number of recent books deal with the theory of groups of linear transformations and its connection with the theory of algebras(1). Most of the work has been restricted to the case of completely reducible systems or, in other words, to semisimple algebras. There are, however, a number of questions which make it desirable not to neglect the other case. The aim of this and a following paper is a study of such not completely reducible systems, in particular of their regular representations. It appeared necessary to start again right from the beginning of the theory, in order to add a number of remarks to well known results and methods(2). The coefficients of the matrices in this paper are taken from an arbitrary division ring K (=skew field or noncommutative field K). This is a generalization of the ordinary theory which does not always work smoothly. For instance, the (left) rank of a ring of matrices $\mathfrak{A}$ is not invariant under similarity transformation. This implies that similar rings $\mathfrak{A}$ and $\mathfrak{A}_1$ may have different regular representations. Yet it is possible to derive a number of results which, in the case of a commutative K, imply the fundamental theorems of Frobenius, Burnside, Loewy, I. Schur and Wedderburn.

Sections 1 and 2 deal with a number of group-theoretical remarks. The first of these are concerned with the Jordan-Hölder theorem. The connection between two composition series is studied more closely, and it is proved that sets of residue systems can be chosen such that they can be used in either composition series. Further, the upper and lower Loewy series of a group are studied. It is shown that the ith factor groups in both have a common constituent. This implies the theorem of Krull and Ore(3) that both series have the same length. In Section 3, the necessary tools from the theory of matrices are described briefly. The following two sections contain an application of the group-theoretical methods to the study of the irreducible and the Loewy constituents of a set of matrices. In Section 6, a number of further remarks are added, for instance a generalization of a theorem of A. H. Clifford(4).

Presented to the Society, April 16, 1938, under the title *On groups of linear transformations*; received by the editors June 17, 1940.

(1) Cf., for instance, Albert [1, 2], Deuring [7], Murnaghan [17], van der Waerden [28, 29], Wedderburn [30], and, in particular, Weyl [31].

(2) For these results and methods, compare the papers given in the bibliography.

(3) Krull [12] proved this for Abelian groups, Ore [22] in the general case.

(4) Clifford [6].

In the second part, the regular representation stands in the foreground and, accordingly, we consider systems $\mathfrak{A}$ of matrices which form semigroups (i.e., are closed under multiplication). There exists a certain reciprocity between $\mathfrak{A}$ and its regular representation $\mathfrak{R}$. In order to show the inner reason for this more clearly, we begin Section 7 with a study of group pairs, first introduced by Pontrjagin(5) in connection with topological investigations. Section 8 deals with the regular representation $\mathfrak{R}$. It is, for instance, shown that $\mathfrak{A}$ and $\mathfrak{R}$ have the same irreducible constituents (except perhaps 0); the number of Loewy constituents in both is either the same or differs by one. A number of further results concerning the distribution of the irreducible parts of the Loewy constituents of $\mathfrak{R}$ are proved.

It now follows that the (left) rank r of an irreducible semigroup $\mathfrak{A}$ is divisible by the degree n. The quotient r/n can be expressed by means of properties of the commuting set (Section 9). This furnishes the basis for the proof of Wedderburn's theorem, and of the generalized Burnside theorem. In Section 10, representations of sets of matrices as direct sums of subsets are studied. Finally, in Sections 11 and 12, rings $\mathfrak{A}$ of matrices of degree a are considered which contain all the scalar multiples kI_a of the unit matrix I_a (k in K). Here, of course, the structure theory of algebras can be obtained in its full extent. It is proved that if $\mathfrak{B}$ is a representation of degree b of $\mathfrak{A}$ then $\mathfrak{B}$ is a constituent of $ab\times\mathfrak{A}$. We are further interested in the connection between the Loewy decomposition of the regular representation, and the structure of the powers of the radical.

We add here a few remarks concerning the notation: The word ring is used for noncommutative rings. We use the expression "l-multiplication" by a ("r-multiplication" by a) in order to express that an element is multiplied on its left side (right side) by a. Except in a few places, there would be no restriction in assuming that the system $\mathfrak{A}$ of matrices forms a ring, but it seems more logical to mention only those properties which are actually needed. Thus $\mathfrak{A}$ can first be any system of matrices, later any semigroup (see above), and in the last section it is assumed to be a ring. The zero-matrix, with any number of rows and columns, is denoted by 0, the unit matrix by I, or more clearly by I_n if n is the degree. Places in matrices or sets of matrices which are left blank are to be filled out with 0-matrices, and stars are used for elements in whose form we are not interested.

1. Remarks on composition series

1. We consider groups $\mathfrak{G}$ with a given set of operators(6) Γ which have a finite composition series

$$(1) \qquad \mathfrak{G} = \mathfrak{G}_0 \supset \mathfrak{G}_1 \supset \cdots \supset \mathfrak{G}_r = \{1\}.$$

(5) Pontrjagin [23].

(6) Cf., for instance, van der Waerden [28, vol. 1, §38]. It is easy to extend the definitions

Let $\mathfrak{H}$ be a second group with the same operators which has a composition series

$$(2) \qquad \mathfrak{H} = \mathfrak{H}_0 \supset \mathfrak{H}_1 \supset \cdots \supset \mathfrak{H}_s = \{1\}.$$

We assume that a homomorphism θ is given which maps $\mathfrak{H}$ upon a normal subgroup $\mathfrak{H}^*$ of $\mathfrak{G}$, $\mathfrak{H}^* \subseteq \mathfrak{G}$.

(1.1A) *We can choose complete residues systems* $\mathfrak{P}_\rho$ *of* $\mathfrak{G}_{\rho-1}$ (mod $\mathfrak{G}_\rho$) *and* $\mathfrak{Q}_\sigma$ *of* $\mathfrak{H}_{\sigma-1}$ (mod $\mathfrak{H}_\sigma$), $(\rho=1, 2, \cdots, r; \sigma=1, 2, \cdots, s)$ *such that* (a) θ *either maps* $\mathfrak{Q}_\sigma$ *on a* $\mathfrak{P}_\rho$ *in a* (1-1) *manner and* $\mathfrak{G}_{\rho-1}/\mathfrak{G}_\rho \simeq \mathfrak{H}_{\sigma-1}/\mathfrak{H}_\sigma$, *or* θ *maps* $\mathfrak{Q}_\sigma$ *on* 1. (b) *Each* $\mathfrak{P}_\rho$ *is the image of at most one* $\mathfrak{Q}_\sigma$.

Proof. We denote by H^* the image on which θ maps the element H of $\mathfrak{H}$. Similarly, let $\mathfrak{K}^*$ be the image of an arbitrary subset $\mathfrak{K}$ of $\mathfrak{H}$. We choose arbitrary residue systems $\mathfrak{Q}_\sigma$ for $\mathfrak{H}_{\sigma-1}$ (mod $\mathfrak{H}_\sigma$) which contain the unit element. Every H in $\mathfrak{H}$ possesses a unique representation

$$(3) \qquad H = Q_1 Q_2 \cdots Q_s, \qquad Q_\sigma \text{ in } \mathfrak{Q}_\sigma;$$

we have $\mathfrak{H}_\sigma = \mathfrak{Q}_{\sigma+1}\mathfrak{Q}_{\sigma+2} \cdots \mathfrak{Q}_s$. If we change $\mathfrak{Q}_\sigma$ by multiplying its elements by elements of $\mathfrak{H}_\sigma$, we can obtain the most general residue system of $\mathfrak{H}_{\sigma-1}$ (mod $\mathfrak{H}_\sigma$). By a succession of such changes, we shall arrive at a set of residue systems for which (1.1A) holds.

We assume that (1.1A) holds for groups $\mathfrak{G}$ which have a shorter composition series than (1). In particular, (1.1A) will be true for $\mathfrak{G}_1$ in place of $\mathfrak{G}$. If $\mathfrak{H}^* \subseteq \mathfrak{G}$, then we may apply (1.1A) to $\mathfrak{G}_1$ and $\mathfrak{H}$ and see that it also holds for $\mathfrak{G}$ and $\mathfrak{H}$; the residue system $\mathfrak{P}_1$ can be taken arbitrarily.

If $\mathfrak{H}^*$ is not a subgroup of $\mathfrak{G}_1$, then $\mathfrak{H}^*\mathfrak{G}_1 = \mathfrak{G}$. Let j be the first integer for which $\mathfrak{H}_j^*\mathfrak{G}_1 \neq \mathfrak{G}$. Then $\mathfrak{H}_j^*\mathfrak{G}_1$ is a proper normal subgroup of $\mathfrak{H}_{j-1}^*\mathfrak{G}_1 = \mathfrak{G}$ which contains $\mathfrak{G}_1$. Hence $\mathfrak{H}_j^*\mathfrak{G}_1 = \mathfrak{G}_1$, i.e., $\mathfrak{H}_j^* \subseteq \mathfrak{G}_1$. We can define a homomorphic mapping of $\mathfrak{H}_{j-1}/\mathfrak{H}_j$ upon $\mathfrak{H}_{j-1}^*\mathfrak{G}_1/\mathfrak{H}_j^*\mathfrak{G}_1 = \mathfrak{G}/\mathfrak{G}_1$ by

$$(4) \qquad H\mathfrak{H}_j \to H^*\mathfrak{H}_j^*\mathfrak{G}_1 = H^*\mathfrak{G}_1, \qquad H \text{ in } \mathfrak{H}_{j-1}.$$

Since $\mathfrak{H}_{j-1}/\mathfrak{H}_j$ is simple, this is an isomorphism. It follows that $\mathfrak{Q}_j^*$ is a complete residue system $\mathfrak{P}_1$ of $\mathfrak{G}$ (mod $\mathfrak{G}_1$).

Thus for every H of $\mathfrak{H}$, the element $(H^{-1})^*$ will lie in some residue class $Q_j^*\mathfrak{G}_1$ with Q_j in $\mathfrak{Q}_j$, and then $(HQ_j)^*$ will lie in $\mathfrak{G}_1$. In particular, we can multiply the elements of $\mathfrak{Q}_\sigma$ $(\sigma=1, 2, \cdots, j-1)$ by such elements of $\mathfrak{Q}_j$ that θ maps the products on elements of $\mathfrak{G}_1$. In this manner, we obtain a new residue system of $\mathfrak{H}_{\sigma-1}$ (mod $\mathfrak{H}_\sigma$) which we shall use instead of $\mathfrak{Q}_\sigma$ and de-

---

to the case that the product of an operator η with a group element G is defined only if G belongs to a subgroup of G which may depend on η. When we have a group with operators, we consider only subgroups which are admissible, and homomorphisms and isomorphisms which are operator-homomorphisms and operator-isomorphisms, without always stating this explicitly. We include the case that Γ is empty, i.e., that G is a group in the ordinary sense.

note by $\mathfrak{Q}_\sigma$ again. We then have $\mathfrak{Q}_\sigma^* \subseteq \mathfrak{G}_1$ $(\sigma \leqq j-1)$. For $\sigma > j$, we have $\mathfrak{Q}_\sigma^* \subseteq \mathfrak{H}_j^* \subseteq \mathfrak{G}_1$. Hence $\mathfrak{Q}_\sigma^* \subseteq \mathfrak{G}_1$, for $\sigma \neq j$.

The elements H of $\mathfrak{H}$, whose image H^* lies in $\mathfrak{G}_1$ form a normal subgroup $\mathfrak{H}'$ of $\mathfrak{H}$. Obviously, $\mathfrak{H}'$ consists of those elements (3) for which $Q_j = 1$. If we set $\mathfrak{H}_\sigma' = [\mathfrak{H}, \mathfrak{H}']$, $\mathfrak{H}_\sigma'$ is obtained from $\mathfrak{H}_\sigma = \mathfrak{Q}_{\sigma+1}\mathfrak{Q}_{\sigma+2} \cdots \mathfrak{Q}_s$ by removing the factor $\mathfrak{Q}_j$ (if it appears). The groups

$$\mathfrak{H}' \supset \mathfrak{H}_1' \supset \cdots \supset \mathfrak{H}_{j-1}' = \mathfrak{H}_j' \supset \mathfrak{H}_{j+1}' \supset \cdots \supset \mathfrak{H}_s' = \{1\}$$

form a composition series, and $\mathfrak{Q}_1, \cdots, \mathfrak{Q}_{j-1}, \mathfrak{Q}_{j+1}, \cdots, \mathfrak{Q}_s$ are a corresponding set of residue systems; $\mathfrak{H}_{\sigma-1}'/\mathfrak{H}_\sigma' \simeq \mathfrak{H}_{\sigma-1}/\mathfrak{H}_\sigma$ for $\sigma \neq j$. Since θ maps $\mathfrak{H}'$ on the normal subgroup $[\mathfrak{H}^*, \mathfrak{G}_1]$ of $\mathfrak{G}_1$, we may apply the statement (1.1A) to the groups $\mathfrak{G}_1$ and $\mathfrak{H}'$ (in place of $\mathfrak{G}$ and $\mathfrak{H}$), in which case it is assumed to be true. We may have to change the residue classes $\mathfrak{Q}_1, \cdots, \mathfrak{Q}_{j-1}, \mathfrak{Q}_{j+1}, \cdots, \mathfrak{Q}_s$ still further by multiplying the elements of $\mathfrak{Q}_\sigma$ by elements of $\mathfrak{H}_\sigma'$. But because $\mathfrak{H}_\sigma' \subseteq \mathfrak{H}_\sigma$, this change is also possible in the set of residue classes belonging to (2). This shows that (1.1A) is correct for $\mathfrak{G}$ and $\mathfrak{H}$(7).

At the same time we see

(1.1B) *The conditions of* (1.1A) *can be satisfied by choosing the elements of each* $\mathfrak{Q}_\sigma$ *from a certain subgroup* $\mathfrak{J}_\sigma$ *of* $\mathfrak{H}$, *and each* $\mathfrak{P}_\rho$ *either as the image of such a* $\mathfrak{Q}_\sigma$ *or as an arbitrary residue system of* $\mathfrak{G}_{\rho-1}$ *modulo* $\mathfrak{G}_\rho$.

2. If $\mathfrak{H}^* = \mathfrak{G}$, every $\mathfrak{P}_\rho$ will appear in the form $\mathfrak{Q}_\sigma^*$. If, on the other hand, the homomorphism θ is an isomorphism, every $\mathfrak{Q}_\sigma^*$ will appear as a $\mathfrak{P}_\rho$. We now take $\mathfrak{G} = \mathfrak{H}$ and θ as the identical mapping. Then (1.1A) gives the Jordan-Hölder theorem and the first part of the following theorem:

(1.2A) *If two composition series of* $\mathfrak{G}$ *are given, the residue systems* $\mathfrak{P}_\rho$ *can be chosen such that they can be used in both composition series* (*in a different arrangement*). *It is possible to carry one arrangement of the* $\mathfrak{P}_\rho$ *into the other one by successively interchanging two consecutive* $\mathfrak{P}_\rho$ *such that each intermediate arrangement also belongs to a composition series of* $\mathfrak{G}$.

In order to prove the second part, we use the same notation as in §1.1. We now have $r = s$, $\mathfrak{P}_1 = \mathfrak{Q}_j$, $\mathfrak{H}' = \mathfrak{G}_1$. The element

$$\overset{-1}{Q}_{j-1}\overset{-1}{Q}_j Q_{j-1} Q_j, \qquad Q_j \text{ in } \mathfrak{Q}_j,\ Q_{j-1} \text{ in } \mathfrak{Q}_{j-1},$$

lies in $\mathfrak{H}'$ and in $\mathfrak{H}_{j-1}$, if $j > 1$. Since $[\mathfrak{H}', \mathfrak{H}_{j-1}] = \mathfrak{H}_j$, it follows that Q_{j-1} and Q_j commute (mod $\mathfrak{H}_j$). If we interchange $\mathfrak{Q}_{j-1}$ and $\mathfrak{Q}_j$ in $\mathfrak{Q}_1, \mathfrak{Q}_2, \cdots, \mathfrak{Q}_s$, we obtain a set of residue systems belonging to the composition series

(7) In a similar manner, we can prove a theorem which has the same relation to Schreier's extension of the Jordan-Hölder theorem (Schreier [25], Zassenhaus [32]) as (1.1A) has to the Jordan-Hölder theorem itself.

$$\mathfrak{H} = \mathfrak{H}_0 \supset \mathfrak{H}_1 \supset \cdots \supset \mathfrak{H}_{j-2} \supset \mathfrak{H}'_{j-2} \supset \mathfrak{H}_j \supset \cdots \supset \mathfrak{H}_s = \{1\}$$

because $\mathfrak{Q}_{j-1}\mathfrak{Q}_j\mathfrak{H}_j = \mathfrak{Q}_j\mathfrak{Q}_{j-1}\mathfrak{H}_j$ and $\mathfrak{H}'_{j-2} = \mathfrak{Q}_{j-1}\mathfrak{Q}_{j+1}\mathfrak{Q}_{j+2}\cdots\mathfrak{Q}_s$. We next interchange $\mathfrak{Q}_j$ with $\mathfrak{Q}_{j-2}$, etc., until $\mathfrak{Q}_j$ finally stands at the first place. If (1.2A) is true for $\mathfrak{G}_1$, as we may assume, it now follows for $\mathfrak{G}$.

2. Loewy series

1. A group $\mathfrak{G}$ is *completely reducible*(8), if it is the direct product of simple groups $\mathfrak{P}_1, \mathfrak{P}_2, \cdots, \mathfrak{P}_r$. As indicated by this notation,

$$\mathfrak{G} = \mathfrak{P}_1\mathfrak{P}_2\cdots\mathfrak{P}_r, \quad \mathfrak{G}_1 = \mathfrak{P}_2\cdots\mathfrak{P}_r, \cdots, \mathfrak{G}_{r-1} = \mathfrak{P}_r, \quad \mathfrak{G}_r = \{1\}$$

is a composition series. Every normal simple subgroup $\mathfrak{M}$ of $\mathfrak{G}$ is completely reducible and is a direct factor, i.e., $\mathfrak{G} = \mathfrak{M} \times \mathfrak{N}$, where $\mathfrak{N}$ is a normal subgroup of $\mathfrak{G}$. Because $\mathfrak{G}/\mathfrak{M} \simeq \mathfrak{N}$, the factor group $\mathfrak{G}/\mathfrak{M}$ is also completely reducible.

If $\mathfrak{A}$ is a normal subgroup of an arbitrary group $\mathfrak{G}$, we say that $\mathfrak{A}$ is *completely reducible with regard to* $\mathfrak{G}$, if $\mathfrak{A}$ is the direct product of minimal normal subgroups of $\mathfrak{G}$. More generally, if $\mathfrak{A}$ and $\mathfrak{B}$ are normal subgroups of $\mathfrak{G}$ and $\mathfrak{A} \supseteq \mathfrak{B}$, we say that $\mathfrak{A}/\mathfrak{B}$ is completely reducible with regard to $\mathfrak{G}$, if $\mathfrak{A}/\mathfrak{B}$ is completely reducible with regard to $\mathfrak{G}/\mathfrak{B}$. If we add the inner automorphism of $\mathfrak{G}$ to the operators of the groups considered (subgroups of $\mathfrak{G}$ and factor groups formed out of them), then complete reducibility of $\mathfrak{A}/\mathfrak{B}$ with regard to $\mathfrak{G}$ means the same as ordinary complete reducibility of $\mathfrak{A}/\mathfrak{B}$. In the case of abelian groups $\mathfrak{G}$, the words "with regard to $\mathfrak{G}$" can always be omitted.

For any group $\mathfrak{G}$, we prove easily:

(2.1A) *If $\mathfrak{L}$ and $\mathfrak{M}$ are normal subgroups of $\mathfrak{G}$ which are completely reducible with regard to $\mathfrak{G}$, the same is true for $\mathfrak{L}\mathfrak{M}$.*

Proof. We add the set of all inner automorphisms of $\mathfrak{G}$ to the set of operators. If $\mathfrak{D} = [\mathfrak{L}, \mathfrak{M}]$, we may set $\mathfrak{L} = \mathfrak{L}_1 \times \mathfrak{D}$, $\mathfrak{M} = \mathfrak{M}_1 \times \mathfrak{D}$, where $\mathfrak{L}_1$ and $\mathfrak{M}_1$ are normal subgroups of $\mathfrak{G}$. We then have $\mathfrak{L}\mathfrak{M} = \mathfrak{L}_1 \times \mathfrak{M}_1 \times \mathfrak{D}$, since $[\mathfrak{L}_1, \mathfrak{M}_1 \times \mathfrak{D}] = \{1\}$. This shows that (2.1A) is true.

(2.1B) *If $\mathfrak{A}$, $\mathfrak{B}$ and $\mathfrak{J}$ are normal subgroups of $\mathfrak{G}$, where $\mathfrak{B} \subseteq \mathfrak{A}$, and $\mathfrak{A}/\mathfrak{B}$ is completely reducible with regard to $\mathfrak{G}$, then $[\mathfrak{J}, \mathfrak{A}]/[\mathfrak{J}, \mathfrak{B}]$ is completely reducible with regard to $\mathfrak{G}$, and isomorphic with a normal subgroup of $\mathfrak{A}/\mathfrak{B}$.*

Proof. We extend the domain of operators as in the proof of (2.1A). The statement is a consequence from the fact that $[\mathfrak{J}, \mathfrak{A}]\mathfrak{B}/\mathfrak{B} \simeq [\mathfrak{J}, \mathfrak{A}]/[\mathfrak{J}, \mathfrak{A}, \mathfrak{B}] = [\mathfrak{J}, \mathfrak{A}]/[\mathfrak{J}, \mathfrak{B}]$, since $[\mathfrak{J}, \mathfrak{A}]\mathfrak{B}/\mathfrak{B}$ is a normal subgroup of $\mathfrak{A}/\mathfrak{B}$.

(2.1C) *If $\mathfrak{B}$ and $\mathfrak{C}$ are normal subgroups of $\mathfrak{G}$, where $\mathfrak{B}\mathfrak{C}/\mathfrak{B}$ and $\mathfrak{B}\mathfrak{C}/\mathfrak{C}$ are both completely reducible with regard to $\mathfrak{G}$, so is $\mathfrak{B}\mathfrak{C}/[\mathfrak{B}, \mathfrak{C}]$.*

Proof. From (2.1B) it follows that $[\mathfrak{B}\mathfrak{C}, \mathfrak{C}]/[\mathfrak{B}, \mathfrak{C}] = \mathfrak{C}/[\mathfrak{B}, \mathfrak{C}]$ is com-

(8) Cf. van der Waerden [28, vol. 1, p. 143].

pletely reducible with regard to $\mathfrak{G}$. The same is true for $\mathfrak{B}/[\mathfrak{B}, \mathfrak{C}]$. Then (2.1A) shows that $\mathfrak{B}\mathfrak{C}/[\mathfrak{B}, \mathfrak{C}]$ is completely reducible with regard to $\mathfrak{G}/[\mathfrak{B}, \mathfrak{C}]$, and hence with regard to $\mathfrak{G}$.

2. A *Loewy series* of $\mathfrak{G}$ is a series of normal subgroups of $\mathfrak{G}$:

$$(5) \qquad \mathfrak{G} = \mathfrak{M}_0 \supset \mathfrak{M}_1 \supset \mathfrak{M}_2 \supset \cdots \supset \mathfrak{M}_{t-1} \supset \mathfrak{M}_t = \{1\}$$

in which each factor group $\mathfrak{M}_{\tau-1}/\mathfrak{M}_\tau$ is completely reducible with regard to $\mathfrak{G}$.

Of special importance is the *lower Loewy series* (or lower cover series of $\mathfrak{G}$). Here $\mathfrak{M}_{t-1}$ is the normal cover ("Sockel")(9) of $\mathfrak{G}$, i.e., the union of all minimal normal subgroups of $\mathfrak{G}$. It follows from (2.1A) that $\mathfrak{M}_{t-1}$ is completely reducible with regard to $\mathfrak{G}$. More generally, we take for $\mathfrak{M}_{\tau-1}$ the group for which $\mathfrak{M}_{\tau-1}/\mathfrak{M}_\tau$ is the normal cover of $\mathfrak{G}/\mathfrak{M}_\tau$ $(\tau=t, t-1, \cdots)$. Then we actually obtain a Loewy series of $\mathfrak{G}$. Obviously, $\mathfrak{M}_{\tau-1}$ is the largest group which can precede $\mathfrak{M}_\tau$ in any Loewy series of $\mathfrak{G}$.

Let $\mathfrak{H}$ be a second group, and

$$(6) \qquad \mathfrak{H} = \mathfrak{N}_0 \supset \mathfrak{N}_1 \supset \mathfrak{N}_2 \supset \cdots \supset \mathfrak{N}_{u-1} \supset \mathfrak{N}_u = \{1\}$$

be a Loewy series of $\mathfrak{H}$. We then state

(2.2A) *Let θ be a homomorphic mapping of $\mathfrak{H}$ upon a subgroup $\mathfrak{H}^*$ of $\mathfrak{G}$ $(\mathfrak{H}^* \subseteq \mathfrak{G})$ which maps normal subgroups $\mathfrak{N}$ upon normal subgroups $\mathfrak{N}^*$ of $\mathfrak{G}$*(10). *If* (5) *is the lower Loewy series of $\mathfrak{G}$, and* (6) *any Loewy series of $\mathfrak{H}$, then*

$$\mathfrak{N}_{u-1}^* \subseteq \mathfrak{M}_{t-1},\ \mathfrak{N}_{u-2}^* \subseteq \mathfrak{M}_{t-2}, \cdots, \mathfrak{N}_{u-\rho}^* \subseteq \mathfrak{M}_{t-\rho}, \cdots.$$

Proof. Let $\mathfrak{N}$ be a minimal normal subgroup of $\mathfrak{H}$. If its image $\mathfrak{N}^*$ contains a normal subgroup $\mathfrak{T}$ of $\mathfrak{G}$ with $\{1\} \subset \mathfrak{T} \subset \mathfrak{N}^*$, the elements of $\mathfrak{N}$ which are mapped upon elements of $\mathfrak{T}$ form a proper subgroup of $\mathfrak{N}$ which is normal in $\mathfrak{H}$. This is impossible, and hence $\mathfrak{N}^*$ is a minimal normal subgroup of $\mathfrak{G}$, and belongs therefore to $\mathfrak{M}_{t-1}$, the normal cover of $\mathfrak{G}$. It now follows easily that $\mathfrak{N}_{u-1}^* \subseteq \mathfrak{M}_{t-1}$. The mapping θ induces a homomorphic mapping of $\mathfrak{H}/\mathfrak{N}_{u-1}$ upon a subgroup of $\mathfrak{G}/\mathfrak{M}_{t-1}$, which maps normal subgroups upon normal subgroups. Using the same argument, we obtain $(\mathfrak{N}_{u-2}/\mathfrak{N}_{u-1})^* \subseteq (\mathfrak{M}_{t-2}/\mathfrak{M}_{t-1})$, and hence $\mathfrak{N}_{u-2}^* \subseteq \mathfrak{M}_{t-2}$, etc.

3. The dual of the lower Loewy series is the *upper Loewy series* or upper cover series. Here, $\mathfrak{M}_\tau$ is the upper cover of $\mathfrak{M}_{\tau-1}$(11), i.e., the intersection of all maximal normal subgroups of $\mathfrak{M}_{\tau-1}$, $\tau=1, 2, \cdots$. We see successively that $\mathfrak{M}_1, \mathfrak{M}_2, \cdots$ are normal in $\mathfrak{G}$. Then $\mathfrak{M}_\tau$ can also be defined as the intersection of the normal subgroups of $\mathfrak{G}$ which are maximal in $\mathfrak{M}_{\tau-1}$. From

(9) Remak [24], Cf. also Ore [22].

(10) This assumption is necessary whereas in the dual theorem (2.3A) it is sufficient to assume that $\mathfrak{H}^*$ is normal in $\mathfrak{G}$.

(11) Ore [22].

(2.1C) it follows easily that $\mathfrak{M}_{r-1}/\mathfrak{M}_r$ is completely reducible with regard to $\mathfrak{G}$, so that we actually have a Loewy series. Obviously, $\mathfrak{M}_r$ is the smallest group which can follow $\mathfrak{M}_{r-1}$ in any Loewy series of $\mathfrak{G}$.

We now show

(2.3A) *Let θ be a homomorphic mapping of $\mathfrak{H}$ upon a normal subgroup $\mathfrak{H}^*$ of $\mathfrak{G}$ ($\mathfrak{H}^* \subseteq \mathfrak{G}$). If* (6) *is the upper Loewy series of $\mathfrak{H}$, and* (5) *any Loewy series of $\mathfrak{G}$, then $\mathfrak{N}_\rho^* \subseteq \mathfrak{M}_\rho$ ($\rho = 1, 2, \cdots$) where $\mathfrak{N}_\rho^*$ again denotes the image of $\mathfrak{N}_\rho$.*

Proof. Without restriction, we may assume that to every inner automorphism of $\mathfrak{H}$ there corresponds an operator in Γ which produces this automorphism. Form

$$\bar{\mathfrak{G}} = [\mathfrak{H}^*, \mathfrak{G}] = \mathfrak{H}^* \supseteq [\mathfrak{H}^*, \mathfrak{M}_1] \supseteq [\mathfrak{H}^*, \mathfrak{M}_2] \supseteq \cdots. \tag{5'}$$

The distinct groups in (5′) form a Loewy series as follows from (2.1B), and θ maps $\mathfrak{H}$ upon $\bar{\mathfrak{G}}$. We replace $\mathfrak{G}$ by $\bar{\mathfrak{G}}$, and (5) by this Loewy series. If we can prove (2.3A) in this case, it also will be true in the original case. It is, therefore, sufficient to prove (2.3A) in the case where $\mathfrak{G} = \mathfrak{H}^*$. Here, $\mathfrak{N}_\rho^*$ is a normal subgroup of $\mathfrak{G}$. The totality of elements of $\mathfrak{H}$ whose images lie in $\mathfrak{M}_1$ form a normal subgroup $\mathfrak{T}$ of $\mathfrak{H}$. We map $\mathfrak{H}/\mathfrak{T}$ upon $\mathfrak{G}/\mathfrak{M}_1$ by $H\mathfrak{T} \rightarrow H^*\mathfrak{M}_1$ (H in $\mathfrak{H}$). Since $H^*\mathfrak{M}_1 = \mathfrak{M}_1$ only if H is in $\mathfrak{T}$, this mapping is an isomorphism. With $\mathfrak{G}/\mathfrak{M}_1$, then $\mathfrak{H}/\mathfrak{T}$ also is completely reducible, and hence $\mathfrak{T}$ contains the upper cover $\mathfrak{N}_1$ of $\mathfrak{H}$. This implies $\mathfrak{N}_1^* \subseteq \mathfrak{M}_1$. If for $\mathfrak{M}_1$, $\mathfrak{N}_1$, and the mapping induced by θ the statement has been proved, as we may assume, it now follows for $\mathfrak{G}$, $\mathfrak{H}$ and the mapping θ.

4. We now consider the case that $\mathfrak{G} = \mathfrak{H}$, and θ is the identical isomorphism. From (2.2A) it follows that any Loewy series (6) of $\mathfrak{G}$ has at least the same length as the lower Loewy series (5), since for $u < t$ we would have $\mathfrak{N}_0^* = \mathfrak{G} \subseteq \mathfrak{M}_{t-u} \subset \mathfrak{M}_0 = \mathfrak{G}$. Similarly, it follows from (2.3A) that any Loewy series has at least the same length as the upper Loewy series. (If we use the notation of (5) and (6) for these series, and if we have $t < u$, then $\{1\} = \mathfrak{M}_t \supseteq \mathfrak{N}_t^* \neq \{1\}$ which is impossible.) If we take for (5) the lower and for (6) the upper Loewy series of $\mathfrak{G} = \mathfrak{H}$, we have $t = u$, hence

(2.4A) *The lower and the upper Loewy series of $\mathfrak{G}$ have the same length*[12].

From (2.3A), we obtain $\mathfrak{N}_\rho \subseteq \mathfrak{M}_\rho$ in our case. If we had $\mathfrak{N}_{\rho-1} \subseteq \mathfrak{M}_\rho$, we could apply (2.3A) to the Loewy series

$$\mathfrak{M}_\rho \supset \mathfrak{M}_{\rho+1} \supset \cdots \supset \mathfrak{M}_u = \{1\}, \quad \mathfrak{N}_{\rho-1} \supset \mathfrak{N}_\rho \supset \cdots \supset \mathfrak{N}_u = \{1\},$$

of which the second one is the upper Loewy series of $\mathfrak{N}_{\rho-1}$. We then find $\mathfrak{N}_\rho \subseteq \mathfrak{M}_{\rho+1}, \cdots, \mathfrak{N}_{u-1} \subseteq \mathfrak{M}_u = \{1\}$ which is impossible. Consequently, $\mathfrak{N}_{\rho-1}$ contains elements which do not belong to $\mathfrak{M}_\rho$, and hence $\mathfrak{N}_\rho \subseteq [\mathfrak{N}_{\rho-1}, \mathfrak{M}_\rho] \subset \mathfrak{N}_{\rho-1}$ and $\mathfrak{M}_\rho \subset \mathfrak{N}_{\rho-1}\mathfrak{M}_\rho \subseteq \mathfrak{M}_{\rho-1}$. Since $\mathfrak{N}_{\rho-1}\mathfrak{M}_\rho/\mathfrak{M}_\rho \simeq \mathfrak{N}_{\rho-1}/[\mathfrak{N}_{\rho-1}, \mathfrak{M}_\rho]$, we obtain

(2.4B) *The ρth factor groups $\mathfrak{M}_{\rho-1}/\mathfrak{M}_\rho$ and $\mathfrak{N}_{\rho-1}/\mathfrak{N}_\rho$ of the lower and upper*

[12] Krull [12], Ore [22], cf.([9]).

Loewy series of $\mathfrak{G}$ contain at least one pair of isomorphic normal subgroups $(\neq\{1\})$.

5. We assume now that $\mathfrak{G}$ is abelian, or, more generally, that all the inner automorphisms of $\mathfrak{G}$ belong to the set of operators. We consider a composition series of $\mathfrak{G}$,

$$\mathfrak{G} = \mathfrak{G}_0 \supset \mathfrak{G}_1 \supset \cdots \supset \mathfrak{G}_r = \{1\},$$

and a corresponding set of residue systems $\mathfrak{P}_1, \mathfrak{P}_2, \cdots, \mathfrak{P}_r$. With regard to a later application, it is desirable to give a method of obtaining the lower Loewy series. It may happen that a $\mathfrak{P}_\rho$ can be chosen to be a (normal([12])) subgroup of $\mathfrak{G}$; we call these $\mathfrak{P}_\rho$ the residue systems of *lowest kind*. We state

(2.5A) *The normal cover of $\mathfrak{G}$ is equal to the product of the residue systems $\mathfrak{P}_\rho$ of lowest kind, if these are chosen to be normal subgroups of $\mathfrak{G}$.*

Proof. It is clear that all these $\mathfrak{P}_\rho$ belong to the normal cover $\mathfrak{H}$ of $\mathfrak{G}$. We determine a set of residue systems $\mathfrak{Q}_1, \mathfrak{Q}_2, \cdots, \mathfrak{Q}_s$ of a composition series of $\mathfrak{H}$ such that each $\mathfrak{Q}_\sigma$ is a minimal normal subgroup of $\mathfrak{G}$ (cf. §2.1), and apply the method of §1.1 to $\mathfrak{G}$, $\mathfrak{H}$, and the identical mapping. If j has the same significance as in §1.1, we may assume that $j=1$, since the $\mathfrak{Q}_\sigma$ here can be permuted arbitrarily. No modification of the $\mathfrak{Q}_\sigma$ is necessary, and one $\mathfrak{P}_\rho$ can be replaced by $\mathfrak{Q}_1$. This shows that this $\mathfrak{P}_\rho$ is of lowest kind. After the next step, one $\mathfrak{P}_\sigma$ will be replaced by $\mathfrak{Q}_2$, etc. Since θ is a (1-1) mapping, every $\mathfrak{Q}_\lambda$ will finally appear. This shows that the number of residue classes of lowest kind cannot be smaller than s. The product of these $\mathfrak{P}_\nu$, chosen as normal subgroups of $\mathfrak{G}$, must give the full normal cover $\mathfrak{M}_{t-1}$ as stated in (2.5A).

We now remove these $\mathfrak{P}_\nu$ of lowest kind from $\mathfrak{P}_1, \mathfrak{P}_2, \cdots, \mathfrak{P}_r$ and work from now on modulo $\mathfrak{M}_{t-1}$. It is clear that the remaining $\mathfrak{P}_\lambda$ form a system of residue classes belonging to a composition series of $\mathfrak{G}/\mathfrak{M}_{t-1}$. Again we single out the residue systems which now are of lowest kind, and choose them such that their elements (mod $\mathfrak{M}_{t-1}$) form normal subgroups of $\mathfrak{G}/\mathfrak{M}_{t-1}$. Their product, multiplied by $\mathfrak{M}_{t-1}$ gives the group $\mathfrak{M}_{t-2}$ in the lower Loewy series. Continuing in this manner, we can obtain this series.

3. Matrices in a division ring

1. There is no difficulty in extending the ordinary theory of matrices to the case in which the coefficients of the matrices are taken from a fixed division ring K (instead of a field). Of course, the products ρA and $A\rho$ of a matrix A and a "scalar" ρ from K will in general be different. Otherwise, there is no difference, as we are not interested in the question of the determinant here. A square matrix M of degree n is nonsingular if there exists a reciprocal M^{-1} with $MM^{-1}=M^{-1}M=I_n$ where $I_n=(\delta_{\kappa\lambda})$, $\delta_{\kappa\kappa}=1$, $\delta_{\kappa\lambda}=0$ for $\kappa\neq\lambda$, is the unit matrix of degree n.

([12]) Any admissible subgroup now is normal.

Let $M_1, M_2, \cdots, M_q$ be matrices of the same type (m, n), i.e., with m rows and n columns. We say that the matrices are *l-independent*, if no linear relation $\alpha_1 M_1 + \cdots + \alpha_q M_q = 0$ exists with coefficients α_κ in K, except for $\alpha_1 = \alpha_2 = \cdots = \alpha_q = 0$. Similarly, the matrices are *r-independent*, if no relation $M_1\alpha_1 + \cdots + M_q\alpha_q = 0$ exists, except for $\alpha_1 = \cdots = \alpha_q = 0$. The *l-rank* of a set $\mathfrak{M}$ of matrices of the same type is defined as the maximum number z of l-independent matrices of $\mathfrak{M}$, and any z such l-independent matrices form an *l-basis* of $\mathfrak{M}$. Correspondingly, the *r-rank* of $\mathfrak{M}$ and *r-basis* of $\mathfrak{M}$ are defined.

2. There is also no difficulty in introducing n-dimensional vector-spaces $\mathfrak{B}$ over a division ring K, and extending the elementary properties of ordinary vector spaces. We arrange the n components x_κ of a vector X with regard to a fixed basis in a column (matrix of type $(n, 1)$). We consider two operations for vectors, addition and r-multiplication with elements of K; these operations appear as a special case of the corresponding operations with matrices. The vector space $\mathfrak{B}$ is an abelian group with addition as group-combination, which possesses the elements of K as operators. It is the direct sum of n simple groups.

We may also consider a second set of vectors U which are given by rows (i.e., matrices of type $(1, n)$). Here we have an addition and an l-multiplication of vectors with elements of K. We denote such vectors as *contragredient* vectors.

A matrix $A = (a_{\kappa\lambda})$ of type (m, n) defines a homomorphic mapping of an n-dimensional vector space upon a subspace of an m-dimensional vector space: $X \rightarrow X^* = AX$, provided that in both spaces coordinate systems have been chosen. The matrix A also defines a homomorphic mapping of an m dimensional contragredient space upon a subspace of an n-dimensional contragredient space: $U \rightarrow U^* = UA$.

3. Let $m = m_1 + m_2 + \cdots + m_k$ and $n = n_1 + n_2 + \cdots + n_l$ be partitions of m and n. We often write matrices A of type (m, n) in the form $(A_{\kappa\lambda})$ where $A_{\kappa\lambda}$ itself is a matrix of type (m_κ, n_λ). We then say that A has been *broken up according to the scheme* $(m_1, \cdots, m_k | n_1, \cdots, n_l)$. If $B = (B_{\kappa\lambda})$ is a matrix of type (n, r) which is broken up according to a scheme $(n_1, \cdots, n_l | r_1, \cdots, r_q)$, then $AB = (\sum_\mu A_{\kappa\mu} B_{\mu\lambda})$, i.e., the product can be formed as if $A_{\kappa\lambda}$ and $B_{\kappa\lambda}$ are scalars, provided that the right-hand side has a meaning. The corresponding fact holds for sums of matrices; here A and B must be broken up according to the same scheme.

We also break up the n-dimensional vector X into an n_1-dimensional vector X_1, an n_2-dimensional vector $X_2, \cdots$, an n_l-dimensional vector X_l. The matrices of the following linear transformations are of importance.

$$\begin{aligned}
T_{ij}(Q)&: \quad X_\kappa^* = X_\kappa \quad (\kappa \neq i), \qquad X_i^* = X_i + QX_j;\\
Z_{ij}&: \quad X_\kappa^* = X_\kappa \quad (\kappa \neq i, j), \qquad X_i^* = X_j, \qquad X_j^* = X_i;\\
W_i(P)&: \quad X_\kappa^* = X_\kappa \quad (\kappa \neq i), \qquad X_i^* = PX_i;
\end{aligned}$$

where Q is a matrix of type (n_i, n_j), and P a nonsingular matrix of degree n_i. We denote by A_c a matrix in which the columns are broken up according to the scheme $(n_1, n_2, \cdots, n_l)$, by A_r a matrix in which the rows have been broken up in this manner, by A a square matrix in which both rows and columns have been broken up in this manner. By combining the corresponding linear transformations, we obtain easily

(3.3A) *The matrix $A_cT_{ij}(Q)$ is obtained from A_c by adding the ith column, r-multiplied by Q, to the jth column; $T_{ij}(Q)^{-1}A_r$ is obtained from A_r by subtracting the jth row, l-multiplied by Q, from the ith row. Finally, $T_{ij}(Q)^{-1}AT_{ij}(Q)$ is obtained from A by performing these two operations successively.*

(3.3B) *The matrix A_cZ_{ij} is obtained from A_c by interchanging the ith and jth column; $Z_{ij}^{-1}A_r$ is obtained from A_r by interchanging the ith and jth row; $Z_{ij}^{-1}AZ_{ij}$ is obtained from A by performing both operations.*

(3.3C) *The matrix $A_cW_i(P)$ is obtained from A_c by r-multiplying the ith column by P; $W_i(P)^{-1}A_r$ is obtained from A_r by l-multiplying the ith row by P^{-1}; and $W_i(P)^{-1}AW_i(P)$ is obtained from A by performing both operations.*

4. The operations in §3.3 can be used in particular if all the numbers n_λ are equal to 1, i.e., if the matrices $A=(a_{\kappa\lambda})$ are taken in their original form. We perform with A a succession of operations of the kind mentioned in (3.3A), (3.3B), (3.3C). This amounts to a succession of l-multiplications and r-multiplications of A by nonsingular square matrices. The new matrix then has the form GAH where G and H are themselves nonsingular square matrices. It can easily be seen that the operations may be chosen such that the new matrix has the form[14]

$$GAH = \begin{pmatrix} I_\rho & 0 \\ 0 & 0 \end{pmatrix}. \tag{7}$$

Here, ρ is an integer, the *rank* of A; and $\rho \leqq m$, $\rho \leqq n$.

We now can discuss the solution of linear equations

$$\sum_{\lambda=1}^{n} a_{\kappa\lambda}x_\lambda = b_\kappa, \qquad \kappa = 1, 2, \cdots, m, \tag{8}$$

or, in matrix form, $AX=B$, where B is an m-dimensional vector. We set $X=HX^*$, $X^*=H^{-1}X$. Then (8) becomes identical with $(GAH)X^*=GB$, in which form it can easily be solved because of (7). In particular, in the homogeneous case $B=0$, we have exactly $n-\rho$ r-independent solutions X of (8). This shows that the rank ρ of A is uniquely determined by A. We may also characterize ρ as the r-rank of the set of vectors B which are obtained from

---

[14] The second row or the second column on the right side may be missing.

(8), if X ranges over all n-dimensional vectors. If the division ring K is replaced by a larger division ring $\overline{K}$, the number ρ remains unchanged, and a complete system of r-independent solutions of the homogeneous equations with regard to K will have the corresponding properties with regard to $\overline{K}$. If (8) has no solution in K, it has no solution in $\overline{K}$.

The "contragredient" equations

$$\sum_{\kappa=1}^{m} u_\kappa a_{\kappa\lambda} = b'_\lambda, \qquad \lambda = 1, 2, \cdots, n,$$

for $u_1, u_2, \cdots, u_m$ can be discussed in a similar manner.

From the characterizations of the rank of a matrix, it follows easily that the rank of a product of matrices is not larger than the rank of either factor.

5. Let us define the transpose A' of a matrix $A = (a_{\kappa\lambda})$ so that the ordinary rule $(A_1A_2)' = A_2'A_1'$ holds for any two matrices whose product is defined. We must take A' not as a matrix with coefficients in K but in the antisymmetric division ring K'. This K' consists of all symbols α' where α is an arbitrary element of K. We have $\alpha' = \beta'$, if and only if $\alpha = \beta$, and we define addition and multiplication by

$$\alpha_1' + \alpha_2' = (\alpha_1 + \alpha_2)'; \qquad \alpha_1'\alpha_2' = (\alpha_2\alpha_1)'.$$

If we now set $A' = (a'_{\lambda\kappa})$ (κ, row-index; λ, column-index), we readily obtain $(A_1A_2)' = A_2'A_1'$.

4. The irreducible constituents of a set of square matrices

1. Consider a set $\mathfrak{Z}$ of elements α of any kind, and a number of sets of matrices $\mathfrak{A}, \mathfrak{B}, \cdots$. We assume that to every α in $\mathfrak{Z}$ there corresponds a matrix A_α in $\mathfrak{A}$, a matrix B_α in $\mathfrak{B}$, etc., such that all the matrices of $\mathfrak{A}, \mathfrak{B}, \cdots$ appear at least once in the form $A_\alpha, B_\alpha, \cdots$ respectively. We then say that $\mathfrak{A}, \mathfrak{B}, \cdots$ are related sets. Equations between related sets $\mathfrak{A}, \mathfrak{B}, \mathfrak{C}, \mathfrak{D}$ such as

$$\mathfrak{A} = \mathfrak{B}, \qquad \mathfrak{A} = \begin{pmatrix} \mathfrak{B} & \\ \mathfrak{C} & \mathfrak{D} \end{pmatrix}, \qquad \mathfrak{A}P = P\mathfrak{B} \text{ (with a fixed matrix } P)$$

indicate that for every α in $\mathfrak{Z}$ the corresponding equations hold:

$$A_\alpha = B_\alpha, \qquad A_\alpha = \begin{pmatrix} B_\alpha & \\ C_\alpha & D_\alpha \end{pmatrix}, \qquad A_\alpha P = PB_\alpha.$$

2. Let $\mathfrak{A}$ be a set of square matrices A of degree n interpreted as linear transformations $X \to X^* = AX$ of an n-dimensional vector space $\mathfrak{V}$. If we introduce new coordinates by a linear transformation $x_\kappa = \sum p_{\kappa\lambda} z_\lambda$, the set $\mathfrak{A}$ is replaced by $P^{-1}\mathfrak{A}P$. These two sets $\mathfrak{A}$ and $P^{-1}\mathfrak{A}P$ are *similar*, $\mathfrak{A} \sim P^{-1}\mathfrak{A}P$; they

are related (with $\mathfrak{Z}=\mathfrak{A}$). Similar sets often are considered as not essentially different.

The vectors X of $\mathfrak{B}$ form an additive abelian group, and we can now introduce two kinds of operators: As the first kind of operator, we take the elements ρ of K, the operation being defined as r-multiplication of X by ρ (as before). As the second kind of operator, we take the elements α of $\mathfrak{Z}$, the operation being defined by $\alpha X=A_\alpha X$.

Let $\mathfrak{B}$ be a second set of matrices which is related to $\mathfrak{A}$, and let $\mathfrak{W}$ be a vector space in which the corresponding linear transformations take place. Then $\mathfrak{A}\sim\mathfrak{B}$, if and only if $\mathfrak{B}$ and $\mathfrak{W}$ are operator-isomorphic (with regard to $\mathfrak{Z}$ and K).

More generally, let us assume that we have an operator-homomorphic mapping of $\mathfrak{W}$ upon an admissible subgroup $\mathfrak{B}_0$ of $\mathfrak{B}$. This mapping is given by a linear transformation $Y\rightarrow X=PY$, (Y in $\mathfrak{W}$, X in $\mathfrak{B}$). The condition for an operator-homomorphism with regard to $\mathfrak{Z}$, then, is $\alpha X=P(\alpha Y)$ for every α in $\mathfrak{Z}$, i.e., $A_\alpha PY=PB_\alpha Y$. Since this must hold for every Y in $\mathfrak{W}$, we find

$$\mathfrak{A}P = P\mathfrak{B}.$$

We then say that P *intertwines* $\mathfrak{A}$ and $\mathfrak{B}$. When $\mathfrak{A}$ and $\mathfrak{B}$ are replaced by similar sets $M^{-1}\mathfrak{A}M$ and $N^{-1}\mathfrak{B}N$, we have

$$(M^{-1}\mathfrak{A}M)(M^{-1}PN) = (M^{-1}PN)(N^{-1}\mathfrak{B}N)\text{[15]}$$

and the matrix $M^{-1}PN$ obviously *takes the place of* P.

3. If the group $\mathfrak{B}$ with the sets of operators $\mathfrak{Z}$, K is simple, then $\mathfrak{A}$ is an *irreducible set*. If $\mathfrak{A}$ is reducible, $\mathfrak{B}$ has an admissible subgroup $\tilde{\mathfrak{B}}$ with $\mathfrak{B}\supset\tilde{\mathfrak{B}}\supset\{0\}$. This $\tilde{\mathfrak{B}}$, then, is a linear subspace which is invariant under the transformations of $\mathfrak{A}$. If we choose the basis of $\mathfrak{B}$ such that the last r basis elements form a basis of $\tilde{\mathfrak{B}}$, then $\mathfrak{A}$ splits in the form

$$\mathfrak{A} = \begin{pmatrix} \mathfrak{M}_1 & \\ \mathfrak{M}_3 & \mathfrak{M}_4 \end{pmatrix} \tag{9}$$

where $\mathfrak{A}$ is broken up according to the scheme $(n-r,\ r\,|\,n-r,\ r)$. Conversely if $\mathfrak{A}$ has this form with regard to a suitable coordinate system, then $\mathfrak{A}$ is reducible. Here $\mathfrak{M}_4$ are the transformations induced by $\mathfrak{A}$ in $\tilde{\mathfrak{B}}$, and $\mathfrak{M}_1$ are the transformations induced by $\mathfrak{A}$ in $\mathfrak{B}/\tilde{\mathfrak{B}}$.

We may interpret the matrices of $\mathfrak{A}$ by means of linear transformations $U\rightarrow U^*=UA$ of a contragredient vector space $\mathfrak{W}$. If $\mathfrak{A}$ splits in the form (9), then $\mathfrak{W}$ has an invariant subspace $\tilde{\mathfrak{W}}$ of $n-r$ dimensions, and the transformations of $\mathfrak{A}$ induce the transformations of $\mathfrak{M}_1$ in $\tilde{\mathfrak{W}}$ and those of $\mathfrak{M}_4$ in $\mathfrak{W}/\tilde{\mathfrak{W}}$, so that the roles of $\mathfrak{M}_1$ and $\mathfrak{M}_4$ are interchanged.

4. We now consider a composition series of $\mathfrak{B}$

---

(15) Schur [26].

$$\mathfrak{V} = \mathfrak{V}_0 \supset \mathfrak{V}_1 \supset \mathfrak{V}_2 \supset \cdots \supset \mathfrak{V}_r = \{0\}.$$

Let $E_\nu^{(i)}$ ($\nu=1, 2, \cdots, a_i$) be a maximal set of vectors of $\mathfrak{V}_{i-1}$ which are r-independent (mod $\mathfrak{V}_i$). Then the totality of all vectors

$$E_1^{(i)} z_1 + E_2^{(i)} z_2 + \cdots + E_{a_i}^{(i)} z_{a_i}, \qquad z_\lambda \text{ in } K, \tag{10}$$

form a complete residue system $\mathfrak{P}_i$ of $\mathfrak{V}_{i-1}$ (mod $\mathfrak{V}_i$). All the vectors $E_\nu^{(i)}$, arranged according to increasing i form a basis of $\mathfrak{V}$, and with regard to this basis, $\mathfrak{A}$ has the form

$$\mathfrak{A} \sim \begin{bmatrix} \mathfrak{A}_1 & & & \\ \mathfrak{A}_{21} & \mathfrak{A}_2 & & \\ \cdot & \cdot & \cdot & \cdot \\ \mathfrak{A}_{r1} & \mathfrak{A}_{r2} & \cdots & \mathfrak{A}_r \end{bmatrix} \tag{11}$$

where $\mathfrak{A}_i$ is an irreducible set of square matrices of degree a_i. These $\mathfrak{A}_i$ are called the *irreducible constituents of* $\mathfrak{A}$.

From Jordan-Hölder's theorem, we obtain at once([16])

(4.4A) *The irreducible constituents of a set $\mathfrak{A}$ of square matrices are uniquely determined apart from their arrangement, if similar sets are considered as equal.*

When we replace the $E_\lambda^{(i)}$ by another basis of $\mathfrak{V}_{i-1}$ (mod $\mathfrak{V}_i$), then $\mathfrak{A}_i$ is replaced by a similar set. We obtain this new form of $\mathfrak{A}$ by a similarity transformation of type (3.3C).

If a formula (11) holds where each $\mathfrak{A}_i$ is a reducible or irreducible set of square matrices of some degree a_i, we say that each $\mathfrak{A}_i$ is a *constituent* of $\mathfrak{A}$. In particular, we call $\mathfrak{A}_1$ a *top constituent* and $\mathfrak{A}_r$ a *bottom constituent*.

Let $\mathfrak{A}$ and $\mathfrak{B}$ again be two related intertwined sets, $\mathfrak{A}P = P\mathfrak{B}$ and $P \neq 0$. We consider again the mapping of $\mathfrak{W}$ upon a certain admissible subgroup $\bar{\mathfrak{V}}$ of $\mathfrak{V}$ which is defined by P. The vectors of $\mathfrak{W}$ which are mapped upon 0 form an admissible subgroup $\bar{\mathfrak{W}}$ of $\mathfrak{W}$, and we have $\bar{\mathfrak{V}} \simeq \mathfrak{W}/\bar{\mathfrak{W}}$. If we use these subgroups in order to split $\mathfrak{A}$ and $\mathfrak{B}$, we have with regard to suitable coordinate systems

$$\mathfrak{A} = \begin{pmatrix} * & 0 \\ * & \mathfrak{U} \end{pmatrix}, \qquad P = \begin{pmatrix} 0 & 0 \\ I & 0 \end{pmatrix}, \qquad \mathfrak{B} = \begin{pmatrix} \mathfrak{U} & 0 \\ * & * \end{pmatrix}.$$

This gives Schur's lemma([17]).

([16]) This simple proof for the uniqueness of the irreducible constituents is due to W. Krull [11].

([17]) I. Schur [26]. Schur's proof is extremely simple. By means of (7), similarity transformations of $\mathfrak{A}$ and $\mathfrak{B}$ are performed such that P assumes the desired form, and then $\mathfrak{A}$ and $\mathfrak{B}$ must have the form given here.

(4.4B) *If two related sets $\mathfrak{A}$ and $\mathfrak{B}$ are intertwined by a matrix $P\neq 0$, then there exists a bottom constituent of $\mathfrak{A}$ which appears as a top constituent of $\mathfrak{B}$. If $\mathfrak{A}$ and $\mathfrak{B}$ are irreducible, then P is nonsingular, and $\mathfrak{A}$ and $\mathfrak{B}$ are similar.*

5. We now apply the results of Section 1. We choose the residue systems of $\mathfrak{V}_{i-1}/\mathfrak{V}_i$ always as in (10), consisting of all linear combinations of some r-independent vectors. From (1.1B) we see that the results of Section 1 remain valid, if we restrict the choice of residue systems by this condition.

Any change of the residue system $\mathfrak{P}_j$ as used in Section 1 can be accomplished by a succession of changes of the following kind: The elements of $\mathfrak{P}_j$ are multiplied by elements of some $\mathfrak{P}_i$, with $i>j$. This now corresponds to replacing $E_\nu^{(j)}$ by $E_\nu^{(j)}+S_\nu^{(i)}$ where each $S_\nu^{(i)}$ is of the form (10). This basis transformation corresponds to the linear transformation $X_\kappa^*=X_\kappa$ for $\kappa\neq i$, $X_i^*=X_i-QX_j$ where the vector X is broken up according to the scheme $(a_1, \cdots, a_k|\, 1)$ and the matrix Q of type (a_i, a_j) is formed by the components z_λ, (10), of the vectors $S_\nu^{(i)}$. This is a transformation $T_{ij}(-Q)=T_{ij}(Q)^{-1}$ (cf. §3.3), and $\mathfrak{A}$ is there replaced by $T_{ij}(Q)^{-1}\mathfrak{A}T_{ij}(Q)$. According to (3.3A), we have to add the ith column in (11), r-multiplied by Q, to the jth column, and the jth row, l-multiplied by $-Q$, to the ith row. Because of $i>j$, the triangular form (11) of $\mathfrak{A}$ is not disturbed:

$$\begin{bmatrix} \cdot & & & & \\ \cdot\ \cdot & & & & \\ \text{——}\ \mathfrak{A}_j & \cdot & & & \\ \downarrow & & \cdot & & \\ \text{——}\ \mathfrak{A}_{ij} & & & \mathfrak{A}_i & \cdot \\ * & | & \leftarrow & | & *\ \cdot \end{bmatrix}.$$

Only the sets $\mathfrak{A}_{i\lambda}$ with $\lambda\leqq j$ and $\mathfrak{A}_{\mu j}$ with $\mu\geqq i$ will be changed. We denote such a special similarity transformation of $\mathfrak{A}$ as an *elementary similarity transformation* of $\mathfrak{A}$. All the $\mathfrak{A}_\kappa$ remain unchanged.

Consider again two related sets of square matrices $\mathfrak{A}$ and $\mathfrak{B}$, operating in the vector spaces $\mathfrak{V}$ and $\mathfrak{W}$ respectively. We assume that both split into irreducible constituents

$$(12)\qquad \mathfrak{A}=\begin{bmatrix} \mathfrak{A}_1 & & \\ * & \mathfrak{A}_2 & \\ \cdot\ \cdot & \cdot\ \cdot & \cdot \\ * & * & \cdots\ \mathfrak{A}_r \end{bmatrix}, \qquad \mathfrak{B}=\begin{bmatrix} \mathfrak{B}_1 & & \\ * & \mathfrak{B}_2 & \\ \cdot\ \cdot & \cdot\ \cdot & \cdot \\ * & * & \cdots\ \mathfrak{B}_s \end{bmatrix},$$

where $\mathfrak{A}_\rho$ has the degree a_ρ and $\mathfrak{B}_\sigma$ has the degree b_σ. If P is an intertwining matrix, we break up P according to the scheme $(a_1, \cdots, a_r|\, b_1, \cdots, b_s)$; say

$P=(P_{\kappa\lambda})$. Then the products $\mathfrak{A}P$ and $P\mathfrak{B}$ can be obtained in the ordinary manner (§3.3). We say, therefore, that the intertwining matrix P has been *broken up in accordance with the splitting of* $\mathfrak{A}$ *and* $\mathfrak{B}$ in (12). Application of (1.1A) to the homomorphic mapping of $\mathfrak{W}\mathfrak{B}$ upon a subgroup of $\mathfrak{B}$ then yields

(4.5A) *Let* $\mathfrak{A}$ *and* $\mathfrak{B}$ *be two related sets of square matrices which split into irreducible constituents* (12), *and let* P *be an intertwining matrix. We can apply to* $\mathfrak{A}$ *and* $\mathfrak{B}$ *a succession of elementary similarity transformations such that the matrix* P^* *which afterwards takes the place of* P (*cf.* §4.2) *contains in each row and each column at most one term not equal to* 0, *if broken up in accordance with the splitting of* $\mathfrak{A}$ *and* $\mathfrak{B}$.

If $P^*=(P^*_{\kappa\lambda})$, then $\mathfrak{A}_\kappa P^*_{\kappa\lambda}=P^*_{\kappa\lambda}\mathfrak{B}_\lambda$ because of this form of P^*. If $P^*_{\kappa\lambda}\neq 0$, then $P_{\kappa\lambda}$ is nonsingular, according to (4.4B). Since for a given λ this may occur for at most one value of κ, after a succession of similarity transformations of type (3.3C), each $P^*_{\kappa\lambda}$ is either 0 or a unit matrix.

Assume now that P is nonsingular so that $\mathfrak{A}$ and $\mathfrak{B}$ are similar. Then every row of P^* must contain one $P^*_{\kappa\lambda}\neq 0$, say for instance $P^*_{1j}\neq 0$. We denote the sets similar to $\mathfrak{A}$ and $\mathfrak{B}$, which we have obtained by $\mathfrak{A}$ and $\mathfrak{B}$ again, and use the notation (12). Then it easily follows from $\mathfrak{A}P^*=P^*\mathfrak{B}$ by forming the first rows of the products that

$$0=\mathfrak{B}_{j1},\ 0=\mathfrak{B}_{j2},\ \cdots,\ 0=\mathfrak{B}_{j,j-1},\ \mathfrak{A}_1=\mathfrak{B}_j.$$

We replace $\mathfrak{B}$ by the similar set $Z^{-1}_{j-1,j}\mathfrak{B}Z_{j-1,j}$ (cf. (3.3B)). Because $\mathfrak{B}_{j,j-1}=0$, the triangular form (12) of $\mathfrak{B}$ is not disturbed,

$$\begin{pmatrix} \cdot & & & \\ \updownarrow & \mathfrak{B}_{j-1} & & \\ & 0 & \mathfrak{B}_j & \\ \cdot & & \leftrightarrow & \cdot \end{pmatrix}.$$

The irreducible constituents of $\mathfrak{B}$ remain the same, only $\mathfrak{B}_{j-1}$ and $\mathfrak{B}_j$ are interchanged. Such a similarity transformation of $\mathfrak{B}$ is an *admissible permutation of rows and columns* which can always be applied, if $\mathfrak{B}_{j,j-1}=0$. According to §4.2, P^* must be replaced by $P^*Z_{j-1,j}$, i.e., the columns $j-1$ and j are to be interchanged (3.3B); but the essential properties of P^* are not destroyed. Similarly, we can interchange $\mathfrak{B}_j$ with $\mathfrak{B}_{j-2}, \mathfrak{B}_{j-3}, \cdots, \mathfrak{B}_1$. The matrix P^{**} which takes the place of P will have the first row $(I, 0, \cdots, 0)$. We now work with the second row of P^{**}. The element $P^{**}_{2k}=I$ in it will not stand in the first column. After a number of further admissible permutations of rows and columns, we may bring it into the second column. Continuing in this manner, we will finally replace P by I. This gives (cf. (1.2A))

(4.5B) *If* $\mathfrak{A}$ *and* $\mathfrak{B}$, (12), *are two similar sets of square matrices which break*

up into irreducible constituents, then it is possible to carry $\mathfrak{B}$ *into* $\mathfrak{A}$ *by a succession of similarity transformations of types* (3.3A), (3.3B), *and* (3.3C)[18].

5. The Loewy constituents

1. A set $\mathfrak{A}$ of square matrices of degree n is *completely reducible*, if the corresponding vector space $\mathfrak{B}$ (with $\mathfrak{A}$ and K as sets of operators) is completely reducible. If we choose the composition series of $\mathfrak{B}$ and the $\mathfrak{P}_i$ as in §2.1, then the formula (11) takes the form

$$\mathfrak{A} \sim \begin{pmatrix} \mathfrak{A}_1 & & & \\ & \mathfrak{A}_2 & & \\ & & \ddots & \\ & & & \mathfrak{A}_r \end{pmatrix}, \qquad \mathfrak{A}_\rho \text{ irreducible,}$$

with zeros above and below the main diagonal. Conversely, if such a formula holds, then $\mathfrak{A}$ is completely reducible.

In the general case, let

$$\mathfrak{B} = \mathfrak{M}_0 \supset \mathfrak{M}_1 \supset \mathfrak{M}_2 \supset \cdots \supset \mathfrak{M}_t = \{0\}$$

be a Loewy series for $\mathfrak{B}$. If we choose the basis of $\mathfrak{B}$ by first taking a maximal set of vectors of $\mathfrak{M}_0$ which are r-independent (mod $\mathfrak{M}_1$), then a maximal set of vectors of $\mathfrak{M}_1$ which are r-independent (mod $\mathfrak{M}_2$), etc., then $\mathfrak{A}$ has the form

$$\mathfrak{A} \sim \begin{pmatrix} \mathfrak{K}_1 & & & \\ * & \mathfrak{K}_2 & & \\ & & \ddots & \\ * & * & \cdots & \mathfrak{K}_t \end{pmatrix} \tag{13}$$

and each $\mathfrak{K}_\lambda$ is completely reducible, since $\mathfrak{M}_{\lambda-1}/\mathfrak{M}_\lambda$ is completely reducible. We say that $\mathfrak{A}$ here appears in a *Loewy form*; every Loewy form of $\mathfrak{A}$ is obtained from a Loewy series of $\mathfrak{B}$. Two Loewy forms are of special importance, the *lower and the upper Loewy form*[19], corresponding to the lower and upper Loewy series of $\mathfrak{B}$, both having the same length (cf. (2.4A)) which will be denoted by $L = L(\mathfrak{A})$. We write them:

$$\mathfrak{A} \sim \begin{pmatrix} \mathfrak{L}_L(\mathfrak{A}) & & & \\ * & \mathfrak{L}_{L-1}(\mathfrak{A}) & & \\ \cdots & \cdots & \cdots & \\ * & * & \cdots & \mathfrak{L}_1(\mathfrak{A}) \end{pmatrix} \sim \begin{pmatrix} \tilde{\mathfrak{L}}_1(\mathfrak{A}) & & & \\ * & \tilde{\mathfrak{L}}_2(\mathfrak{A}) & & \\ \cdots & \cdots & \cdots & \\ * & * & \cdots & \tilde{\mathfrak{L}}_L(\mathfrak{A}) \end{pmatrix}, \tag{14}$$

(18) These transformations are to be applied to the form (12) of $\mathfrak{A}$ and $\mathfrak{B}$.
(19) Cf. A. Loewy [14, 15], W. Krull [12], B. L. van der Waerden [29].

where the first is the lower and the second is the upper Loewy form. The *lower Loewy constituents* $\mathfrak{L}_1(\mathfrak{A}), \mathfrak{L}_2(\mathfrak{A}), \cdots$ are numerated starting from the bottom, and the upper Loewy constituents $\tilde{\mathfrak{L}}_1(\mathfrak{A}), \tilde{\mathfrak{L}}_2(\mathfrak{A}), \cdots$ starting from the top.

The constituent $\mathfrak{L}_1(\mathfrak{A})$ is the maximal completely reducible set which can appear as bottom constituent of $\mathfrak{A}$. If $\mathfrak{A}$ splits into $\mathfrak{B}$ and $\mathfrak{L}_1(\mathfrak{A})$, then

$$\mathfrak{L}_{i+1}(\mathfrak{A}) \sim \mathfrak{L}_i(\mathfrak{B}).$$

Similarly, $\tilde{\mathfrak{L}}_1(\mathfrak{A})$ is the maximal completely reducible set which can appear as a top constituent of $\mathfrak{A}$; and if $\mathfrak{A}$ splits into $\tilde{\mathfrak{L}}_1(\mathfrak{A})$ and $\tilde{\mathfrak{B}}$, then

$$\tilde{\mathfrak{L}}_{i+1}(\mathfrak{A}) \sim \tilde{\mathfrak{L}}_i(\tilde{\mathfrak{B}}).$$

The transformations of $\mathfrak{A}$ transform the space $\mathfrak{M}_{i-1}/\mathfrak{M}_j$, $(j \geqq i)$, into a part of itself and induce, therefore, a set of linear transformations in the space. This set is obtained from (13) by removing the rows and columns with an index less than i or greater than j. We denote this set by $\mathfrak{K}(i \cdots j)$; its main diagonal starts with $\mathfrak{K}_i$ and ends with $\mathfrak{K}_j$. Since in the case of the lower and the upper Loewy series the groups $\mathfrak{M}_h$ are uniquely determined, we have

(5.1A) *The constituent $\mathfrak{L}(i \cdots j)$ of the lower Loewy normal form is uniquely determined apart from similarity transformation. The corresponding fact holds for the upper Loewy normal form.*

From (2.4B), we obtain

(5.1B) *The Loewy constituents $\mathfrak{L}_i(\mathfrak{A})$ and $\tilde{\mathfrak{L}}_{L-i}(\mathfrak{A})$ $(L = L(\mathfrak{A}); i = 1, 2, \cdots, L)$ have at least one common irreducible constituent.*

2. Application of the theorems (2.2A) and (2.3A) gives

(5.2A) *Let $\mathfrak{A}$ and $\mathfrak{B}$ be two related sets of square matrices, both written in Loewy form*

$$\mathfrak{A} = \begin{pmatrix} \mathfrak{J}_1 & & & \\ * & \mathfrak{J}_2 & & \\ \cdot & \cdot & \cdot & \\ * & * & \cdots & \mathfrak{J}_s \end{pmatrix}, \quad \mathfrak{B} = \begin{pmatrix} \mathfrak{K}_1 & & & \\ * & \mathfrak{K}_2 & & \\ \cdot & \cdot & \cdot & \\ * & * & \cdots & \mathfrak{K}_t \end{pmatrix} \tag{15}$$

($\mathfrak{J}_\sigma$ and $\mathfrak{K}_\tau$ completely reducible). Let $P = (P_{\kappa\lambda})$ be an intertwining matrix broken up in accordance with the splitting (15) *of $\mathfrak{A}$ and $\mathfrak{B}$ (cf. §4.5). (α) If $\mathfrak{A}$ is in its lower Loewy normal form, then $P_{\kappa\lambda} = 0$ for $s - \kappa > t - \lambda$. (β) If $\mathfrak{B}$ is in its upper Loewy normal form, then $P_{\kappa\lambda} = 0$ for $\kappa < \lambda$.*

In other words: In the case (α), P has the form given in (16α) below; if $s > t$, the first $s - t$ rows in P consist of zeros. In the case (β), P has the form

(16β); for $s<t$, the last $t-s$ columns consist of zeros:

$$(16\alpha)\qquad P = \begin{bmatrix} \cdot & \cdot & \cdot & \cdot \\ \cdot & P_{s-2,t-2} & 0 & 0 \\ \cdot & P_{s-1,t-2} & P_{s-1,t-1} & 0 \\ \cdot & P_{s,t-2} & P_{s,t-1} & P_{s,t} \end{bmatrix},$$

$$(16\beta)\qquad P = \begin{bmatrix} P_{11} & 0 & 0 & \cdot \\ P_{21} & P_{22} & 0 & \cdot \\ P_{31} & P_{32} & P_{33} & \cdot \\ \cdot & \cdot & \cdot & \cdot \end{bmatrix}.$$

3. When a set $\mathfrak{A}$ is given in the form (11), splitting into irreducible constituents, we can use the method of §2.5 in order to determine the Loewy constituents $\mathfrak{L}(\mathfrak{A})$. We consider one constituent $\mathfrak{A}_i$ in $\mathfrak{A}$,

$$(17)\qquad \mathfrak{A} = \begin{bmatrix} * & & \\ * & \mathfrak{A}_i & \\ * & \mathfrak{C} & \mathfrak{D} \end{bmatrix},$$

where the rows and columns $i+1, i+2, \cdots, r$ of (11) are grouped together in $\mathfrak{D}$. If (11) belongs to the composition series $\mathfrak{B}, \mathfrak{B}_1, \mathfrak{B}_2, \cdots, \mathfrak{B}_r$ and $\mathfrak{P}_r$ is a complete residue system of $\mathfrak{B}_{r-1}$ (mod $\mathfrak{B}_r$), then the question is whether we can change $\mathfrak{P}_i$ so that it forms an admissible subgroup. The only freedom which we have is that we can add arbitrary vectors of $\mathfrak{B}_i$ to the basis elements of $\mathfrak{P}_i$. This amounts to an elementary similarity transformation of (17), involving the second and third row and column (cf. §4.5). If after the change $\mathfrak{P}_i$ is an admissible subgroup, then $\mathfrak{C}$ must become 0, since the modified $\mathfrak{P}_i$ are invariant under $\mathfrak{A}$. But an elementary similarity transformation replaces $\mathfrak{C}$ by $\mathfrak{C}+\mathfrak{D}Q-Q\mathfrak{A}_i$; so that the residue system $\mathfrak{P}_i$ will be of the lowest kind, if and only if this is 0 for a suitable Q, and $\mathfrak{A}_i$ will belong to $L_1(\mathfrak{A})$. Hence

(5.3A) *The first Loewy constituent* $\mathfrak{L}_1(\mathfrak{A})$ *consists of those irreducible constituents* $\mathfrak{A}_i$, (15), *for which a matrix* Q *can be determined such that in* (17) $\mathfrak{C}=Q\mathfrak{A}_i-\mathfrak{D}Q$.

After similarity transformations, we may assume that all $\mathfrak{A}_i$ of this type stand in columns in which otherwise only zeros appear. In order to find $\mathfrak{L}_2(\mathfrak{A})$ we have to remove the rows and columns of the $\mathfrak{A}_i$ "of lowest kind" from $\mathfrak{A}$, and treat the remaining set $\mathfrak{B}$ in the same manner; we have $\mathfrak{L}_{r+1}(\mathfrak{A})=\mathfrak{L}_r(\mathfrak{B})$.

Moving all the constituents $\mathfrak{A}$ of lowest kind to the bottom by admissible permutations §4.5, $\mathfrak{L}_1(\mathfrak{A})$ will appear at the bottom of $\mathfrak{A}$. After removing its rows and columns from $\mathfrak{A}$ and treating the remainder in the same fashion, we

finally arrive at the lower Loewy form of $\mathfrak{A}$. It is remarkable in this connection that the criterion (5.3A) only depends on the solution of linear equations for the coefficients of the matrix Q.

4. The dualism between the upper and lower Loewy form can be realized in the following manner. We replace every matrix A of $\mathfrak{A}$ by its transposed A', §3.5. If $\mathfrak{A}$ is in its lower normal form, (14), the new set $\mathfrak{A}'$ formed by all A' will have the following form

$$\mathfrak{A}' = \begin{pmatrix} \mathfrak{L}_L(\mathfrak{A})' & & * \\ & \ddots & \\ & & \mathfrak{L}_1(\mathfrak{A})' \end{pmatrix}.$$

If we arrange the rows and columns in reverse order, $\mathfrak{A}'$ splits into the constituents $\mathfrak{L}_1(\mathfrak{A})', \cdots, \mathfrak{L}_L(\mathfrak{A})'$. In this manner, we easily see that

$$\text{(5.4A)} \qquad \mathfrak{L}_\nu(\mathfrak{A})' = \bar{\mathfrak{L}}_\nu(\mathfrak{A}') \quad (\nu = 1, 2, \cdots, L;\ L = L(\mathfrak{A}) = L(\mathfrak{A}')).$$

Using this method, we can derive results concerning the upper Loewy form from those concerning the lower Loewy form in §5.3.

6. Additional remarks

1. We consider two related sets $\mathfrak{A}$ and $\mathfrak{B}$ of matrices which split completely into irreducible constituents, i.e.,

$$\mathfrak{A} = \begin{pmatrix} \mathfrak{A}_1 & & & \\ & \mathfrak{A}_2 & & \\ & & \ddots & \\ & & & \mathfrak{A}_r \end{pmatrix}, \qquad \mathfrak{B} = \begin{pmatrix} \mathfrak{B}_1 & & & \\ & \mathfrak{B}_2 & & \\ & & \ddots & \\ & & & \mathfrak{B}_s \end{pmatrix}.$$

If P is an intertwining matrix, $\mathfrak{A}P = P\mathfrak{B}$, we break up P according to this splitting, $P = (P_{\kappa\lambda})$ (cf. §4.5). The condition for $P_{\kappa\lambda}$ becomes $\mathfrak{A}_\kappa P_{\kappa\lambda} = P_{\kappa\lambda}\mathfrak{B}_\lambda$. Using Schur's lemma, we obtain

(6.1A) *Let $\mathfrak{A}$ and $\mathfrak{B}$ be two related sets of matrices which split completely into irreducible constituents $\mathfrak{A}_1, \mathfrak{A}_2, \cdots, \mathfrak{A}_r$ and $\mathfrak{B}_1, \mathfrak{B}_2, \cdots, \mathfrak{B}_s$ respectively. If $P = (P_{\kappa\lambda})$ is an intertwining matrix broken up in accordance with the splitting of $\mathfrak{A}$ and $\mathfrak{B}$, then either $P_{\kappa\lambda} = 0$, or $\mathfrak{A}_\kappa \sim \mathfrak{B}_\lambda$ and $P_{\kappa\lambda}$ is nonsingular and intertwines $\mathfrak{A}_\kappa$ and $\mathfrak{B}_\lambda$. Conversely, if these conditions are satisfied $P = (P_{\kappa\lambda})$ intertwines $\mathfrak{A}$ and $\mathfrak{B}$.*

2. The matrices P which intertwine a set $\mathfrak{A}$ of square matrices with itself, $\mathfrak{A}P = P\mathfrak{A}$, form a ring, the *commuting ring* $\mathfrak{C}(\mathfrak{A})$ of $\mathfrak{A}$. If P in $\mathfrak{C}(\mathfrak{A})$ is a nonsingular matrix, then P^{-1} also belongs to $\mathfrak{C}(\mathfrak{A})$. From Schur's lemma, we find that

(6.2A) *The commuting ring of an irreducible set is a division ring.*

Denote by $k\times\mathfrak{A}$ the set which splits completely into k equal constituents $\mathfrak{A}$, and by $[\mathfrak{A}]_k$ the set of all matrices $(A_{\kappa\lambda})$ of degree k in which the $A_{\kappa\lambda}$ are arbitrary elements of $\mathfrak{A}$. I.e.,

$$k \times \mathfrak{A} = \begin{pmatrix} \mathfrak{A} & & & \\ & \mathfrak{A} & & \\ & & \ddots & \\ & & & \mathfrak{A} \end{pmatrix} \text{ (k times)}, \quad [\mathfrak{A}]_k\text{: all } \begin{pmatrix} A_{11} & \cdots & A_{1k} \\ \cdot & \cdots & \cdot \\ A_{k1} & \cdots & A_{kk} \end{pmatrix} \text{ with } A_{\kappa\lambda} \text{ in } \mathfrak{A}.$$

We then state

(6.2B) (α) $\mathfrak{C}(k\times\mathfrak{A})=[\mathfrak{C}(\mathfrak{A})]_k$. ($\beta$) *If $\mathfrak{A}$ contains* 0 *and* I, *then* $\mathfrak{C}([\mathfrak{A}]_k)=k\times\mathfrak{C}(\mathfrak{A})$. ($\gamma$) $\mathfrak{C}(\mathfrak{C}(k\times\mathfrak{A}))=k\times\mathfrak{C}\mathfrak{C}(\mathfrak{A})$.

Proof. (α) follows at once from (6.1A). In the case of (β), let P be a matrix of $\mathfrak{C}([\mathfrak{A}]_k)$ and set $P=(P_{\kappa\lambda})$ where all the $P_{\kappa\lambda}$ have the degree a of $\mathfrak{A}$. We first choose all $A_{\kappa\lambda}=0$ except one, say $A_{\rho\sigma}$. From $(A_{\kappa\lambda})(P_{\kappa\lambda})=(P_{\kappa\lambda})(A_{\kappa\lambda})$, it follows that $A_{\rho\sigma}P_{\sigma\lambda}=0$ for $\lambda\neq\sigma$, $A_{\rho\sigma}P_{\sigma\sigma}=P_{\rho\rho}A_{\rho\sigma}$. Taking first $A_{\rho\sigma}=I_r$, and then taking $\rho=\sigma$ and taking $A_{\rho\rho}$ arbitrarily, we obtain (β). The statement (γ) is obtained from (α) by applying (β) to $\mathfrak{C}(\mathfrak{A})$ instead of $\mathfrak{A}$; the matrices 0 and I belong to $\mathfrak{C}(\mathfrak{A})$.

From (6.1A) and (6.2Bα) also follows

(6.2C) *If $\mathfrak{A}$ splits completely into $k_1\times\mathfrak{A}_1, \cdots, k_r\times\mathfrak{A}_r$, where $\mathfrak{A}_1, \mathfrak{A}_2, \cdots, \mathfrak{A}_r$ are irreducible and not similar, then $\mathfrak{C}(\mathfrak{A})$ splits completely into* $[\mathfrak{C}(\mathfrak{A}_1)]_{k_1}$, $[\mathfrak{C}(\mathfrak{A}_2)]_{k_2}, \cdots, [\mathfrak{C}(\mathfrak{A}_r)]_{k_r}$.

The $\mathfrak{C}(\mathfrak{A}_\kappa)$ here may be reducible or irreducible (see §9.3 below).

In the general case, a structure theory of the ring $\mathfrak{C}(\mathfrak{A})$ is contained as a special case in the results of Fitting[20].

3. With regard to $[\mathfrak{A}]_k$, we can prove

(6.3A) *If $\mathfrak{A}$ is reducible, so is $[\mathfrak{A}]_k$. If $\mathfrak{A}$ is irreducible and contains* 0 *without consisting of the zero matrix, then $[\mathfrak{A}]_k$ is irreducible.*

Proof. If $\mathfrak{A}$ is reducible, we may assume that it splits into two constituents, i.e.,

$$\mathfrak{A} = \begin{pmatrix} \mathfrak{R}_1 & \\ \mathfrak{R}_3 & \mathfrak{R}_4 \end{pmatrix}.$$

Writing every $A_{\kappa\lambda}$ in the corresponding form, $(A_{\kappa\lambda})$ appears as a matrix of degree $2k$. We rearrange the rows and columns, first taking those with an odd index and then those with an even index. After this similarity transformation, $[\mathfrak{A}]_k$ will split.

---

(20) Fitting [8].

If $\mathfrak{A}$ satisfies the assumptions of the second part of (6.3A), and if $[\mathfrak{A}]_k$ were reducible, then $[\mathfrak{T}]_k$ also would be reducible, where $\mathfrak{T}$ is the ring generated by $\mathfrak{A}$. That this is not so can be easily seen from a simple argument of Weyl([21]).

4. Next, we prove an extension of a theorem of A. H. Clifford([22])

(6.4A) *Let $\mathfrak{B}$ be a set of matrices of degree b and denote by $\mathfrak{H}$ the set of all matrices P of degree b for which $\mathfrak{B}P$ and $P\mathfrak{B}$ consist of the same matrices. The total number of irreducible constituents of $\mathfrak{H}$ is at least equal to the number $L(\mathfrak{B})$ of Loewy constituents of $\mathfrak{B}$, §5.1.*

Proof. After a similarity transformation of $\mathfrak{B}$, we may assume that $\mathfrak{B}$ appears in its lower Loewy normal form. Let P be a fixed element of $\mathfrak{H}$. We form the set $\mathfrak{Z}$ of all pairs (B_1, B_2) of two elements B_1, B_2 of $\mathfrak{B}$ for which $B_1P=PB_2$. To every element of $\mathfrak{Z}$ there corresponds a first matrix B_1 and a second matrix B_2. We thus obtain two related sets $\mathfrak{B}_1$ and $\mathfrak{B}_2$ such that $\mathfrak{B}_1P=P\mathfrak{B}_2$. Since $\mathfrak{B}_1$ and $\mathfrak{B}_2$ both consist of the same matrices as $\mathfrak{B}$, both are in their Loewy normal form. We can now apply (5.2A). Since $\mathfrak{B}_1$ and $\mathfrak{B}_2$ both have $L(\mathfrak{B})$ Loewy constituents, it follows that P breaks up into $L(\mathfrak{B})$ constituents the degrees of which are the degrees of the Loewy constituents $\mathfrak{L}_L(\mathfrak{B}), \cdots, \mathfrak{L}_1(\mathfrak{B})$. This holds for every P in $\mathfrak{H}$, and hence for $\mathfrak{H}$.

Clifford's case is obtained by taking for $\mathfrak{B}$ a normal subgroup of an irreducible group $\mathfrak{G}$ of matrices. Here $\mathfrak{H}\supseteq\mathfrak{G}$ and hence $\mathfrak{H}$ is irreducible. Then (6.4A) shows that $L(\mathfrak{B})=1$, i.e., $\mathfrak{B}$ is completely reducible.

If $\mathfrak{A}$ is an irreducible set, we may apply (6.4A) to $\mathfrak{B}=\mathfrak{C}(\mathfrak{A})$. Then $\mathfrak{H}\supseteq\mathfrak{A}$, and hence $\mathfrak{H}$ again is irreducible and $L(\mathfrak{B})=1$, i.e., $\mathfrak{C}(\mathfrak{A})$ is completely reducible. If $\mathfrak{C}(\mathfrak{A})$ had two nonsimilar irreducible constituents, then $\mathfrak{C}(\mathfrak{C}(\mathfrak{A}))$ would be reducible according to (6.2C), and hence $\mathfrak{A}\subseteq\mathfrak{C}(\mathfrak{C}(\mathfrak{A}))$ would be reducible. This gives

(6.4B) *If $\mathfrak{A}$ is an irreducible set of matrices, $\mathfrak{C}(\mathfrak{A})$ is completely reducible, and all its irreducible constituents are similar.*

From (6.2C), we also obtain

(6.4C) *If $\mathfrak{A}$ is completely reducible, so is $\mathfrak{C}(\mathfrak{A})$.*

5. For the actual construction of intertwining matrices, the following remark is sometimes useful.

(6.5A) *Let $\mathfrak{A}$ and $\mathfrak{B}$ be related sets of matrices and assume that $\mathfrak{A}$ consists*

---

(21) Cf. Weyl [31, p. 86]. The basis of the argument is the following remark. If $\mathfrak{A}\neq\{0\}$ is an irreducible semigroup of matrices of degree a, if $Z\neq 0$ is a fixed a-dimensional vector, then every a-dimensional vector can be written as a finite sum $\sum AZc_A$ where the A are elements of $\mathfrak{A}$ and the c_A are elements of K. If this were not so, the vectors of this type would form an invariant subspace.

(22) Clifford [6].

of nonsingular matrices. For corresponding matrices A and B, let the vector U undergo the transformation contragredient to A, and let X undergo the transformation B; i.e.,

$$U \to U^* = UA^{-1}, \qquad X \to X^* = BX. \tag{18}$$

The matrix P intertwines $\mathfrak{A}$ and $\mathfrak{B}$, if and only if UPX is an invariant for each pair of corresponding transformations (18).

Indeed, from $U^*PX^* = UPX$, it follows that $UA^{-1}PBX = UPX$ for all U and X, and hence $A^{-1}PB = P$.

6. We conclude this section by proving some properties of the Loewy constituents of reducible sets.

(6.6A) *If $\mathfrak{A}$ is a reducible set of matrices*

$$\mathfrak{A} \sim \begin{pmatrix} \mathfrak{G} & \\ \mathfrak{Z} & \mathfrak{H} \end{pmatrix}, \tag{19}$$

then $\mathfrak{L}_i(\mathfrak{A})$ splits into $\mathfrak{L}_i(\mathfrak{H})$ and constituents of $\mathfrak{L}_1(\mathfrak{G}), \mathfrak{L}_2(\mathfrak{G}), \cdots, \mathfrak{L}_i(\mathfrak{G})$ ([22]). *Similarly, $\tilde{\mathfrak{L}}_i(\mathfrak{A})$ splits into $\tilde{\mathfrak{L}}_i(\mathfrak{G})$ and constituents of $\tilde{\mathfrak{L}}_1(\mathfrak{H}), \tilde{\mathfrak{L}}_2(\mathfrak{H}), \cdots, \tilde{\mathfrak{L}}_i(\mathfrak{H})$* ([23]).

Proof. We may assume that $\mathfrak{G}$ and $\mathfrak{H}$ both appear in their lower Loewy normal forms. In order to find $\mathfrak{L}_1(\mathfrak{A})$, we may use the method of §5.3. It is obvious that $\mathfrak{L}_1(\mathfrak{A})$ will be built up from $\mathfrak{L}_1(\mathfrak{H})$ and, perhaps, some constituents of $\mathfrak{L}_1(\mathfrak{G})$. We may assume that all these constituents stand in columns which otherwise consist of zeros. Removing the rows and columns of these constituents from $\mathfrak{A}$, we obtain a set

$$\mathfrak{A}^* = \begin{pmatrix} \mathfrak{G}^* & \\ \mathfrak{Z}^* & \mathfrak{H}^* \end{pmatrix}$$

where $\mathfrak{G}^*$ is a top constituent of $\mathfrak{G}$, and $\mathfrak{H}^*$ a top constituent of $\mathfrak{H}$. It is easily seen, using the same method, that if an irreducible constituent of $\mathfrak{G}^*$ belongs to $L_j(\mathfrak{G}^*)$, it belongs in $\mathfrak{G}$ either to $L_j(\mathfrak{G})$ or $L_{j+1}(\mathfrak{G})$. If for $\mathfrak{A}^*$ the first part of the statement has been proved, as we may assume, it follows easily for $\mathfrak{G}$. The second part is obtained from the first by going over to the transposed matrix as in §5.4.

As a corollary:

(6.6B) *We have $L(\mathfrak{A}) \geq L(\mathfrak{G})$ and $L(\mathfrak{A}) \geq L(\mathfrak{H})$.*

The situation is far simpler, if $\mathfrak{Z} = 0$ in (19) of (6.6A). We then have the following:

(6.6C) *If the set $\mathfrak{A}$ breaks up completely into two constituents $\mathfrak{G}$ and $\mathfrak{H}$, then*

([22]) Some of these constituents may be missing.

$L_i(\mathfrak{A})$ breaks up into $L_i(\mathfrak{G})$ and $L_i(\mathfrak{H})$; $\bar{L}_i(\mathfrak{A})$ breaks up into $\bar{L}_i(\mathfrak{G})$ and $\bar{L}_i(\mathfrak{H})$(24). *Further,* $L(\mathfrak{A}) = \max(L(\mathfrak{G}), L(\mathfrak{H}))$.

The proof again is obtained by the method of §5.3 and is similar to, but simpler than that of (6.6A).

7. Group pairs and associated sets of matrices

1. Consider three Abelian groups $\mathfrak{U}$, $\mathfrak{V}$, and $\mathfrak{W}$, each written with addition as group combination. We assume that the "product" uv of an element u of $\mathfrak{U}$ with an element v of $\mathfrak{V}$ is defined as an element of $\mathfrak{W}$ such that the distributive laws hold,

$$(u_1 + u_2)v = u_1v + u_2v, \qquad u(v_1 + v_2) = uv_1 + uv_2,$$

for any u, u_1, u_2 in $\mathfrak{U}$ and any v, v_1, v_2 in $\mathfrak{V}$(25).

If $\mathfrak{U}$ has a set of operators Γ, and $\mathfrak{V}$ a set of operators Δ, we write the operation in $\mathfrak{U}$ as l-multiplication and the operation in $\mathfrak{V}$ as r-multiplication. We then assume that $\mathfrak{W}$ possesses the two sets of operators Γ and Δ, the first corresponding to l-multiplication and the second to r-multiplication, and that the associative laws hold,

$$\gamma(uv) = (\gamma u)v, \qquad (uv)\delta = u(v\delta), \qquad \gamma(w\delta) = (\gamma w)\delta,$$

for any u in $\mathfrak{U}$, v in $\mathfrak{V}$, w in $\mathfrak{W}$, γ in Γ, δ in Δ. If all these conditions are satisfied, we say that $(\mathfrak{U}, \mathfrak{V})$ is a *group pair.*

An *r-annihilator* v_0 is an element of $\mathfrak{V}$ for which $\mathfrak{U}v_0 = 0$, i.e., uv_0 is the zero-element of $\mathfrak{W}$ for every u in $\mathfrak{U}$. All these r-annihilators form an (admissible) subgroup $\mathfrak{V}_0$ of $\mathfrak{V}$. Similarly, the l-annihilators u_0 in $\mathfrak{U}$ with $u_0\mathfrak{V} = 0$ form a subgroup $\mathfrak{U}_0$ of $\mathfrak{U}$. If we set $(\mathfrak{U}_0 + u)(\mathfrak{V}_0 + v) = uv$, then $(\mathfrak{U}/\mathfrak{U}_0, \mathfrak{V}/\mathfrak{V}_0)$ becomes a group pair in which there are no l-annihilators or r-annihilators except the zero elements. Such a group pair is said to be *primitive*(26).

2. Let $(\mathfrak{U}, \mathfrak{V})$ be a group pair in which the zero element is the only l-annihilator: $\mathfrak{U}_0 = 0$. We consider a set $\mathfrak{B}$ of homomorphic(27) mappings B of $\mathfrak{V}$ upon itself or a subgroup of $\mathfrak{V}$. We say that the group pair $(\mathfrak{U}, \mathfrak{V})$ admits the transformations B of $\mathfrak{B}$, if to each $B: v \to v^*$ there corresponds a transformation $A: u \to u^*$ of $\mathfrak{U}$ upon itself or a subgroup of $\mathfrak{U}$, such that

$$u^*v = uv^* \tag{20}$$

for all u in $\mathfrak{U}$ and all v in $\mathfrak{V}$. The element u^* is uniquely determined by (20), if B and u are given. Further

(7.2A) *The mapping A is a homomorphism.*

---

(24) Some of these constituents may be missing.

(25) Such group pairs $\mathfrak{U}$, $\mathfrak{V}$ have first been considered by Pontrjagin [23].

(26) Cf. Pontrjagin [23].

(27) As always, this is to mean operator-homomorphic mappings.

Proof. We have (for u, u_1, u_2 in $\mathfrak{U}$, v in $\mathfrak{V}$, γ in Γ)

$$(u_1 + u_2)^*v = (u_1 + u_2)v^* = u_1v^* + u_2v^* = u_1^*v + u_2^*v = (u_1^* + u_2^*)v,$$
$$(\gamma u)^*v = (\gamma u)v^* = \gamma(uv^*) = \gamma(u^*v) = (\gamma u^*)v$$

which imply $(u_1+u_2)^* = u_1^*+u_2^*$, $(\gamma u)^* = \gamma u^*$.

We call the set $\mathfrak{A}$ of all these transformations A *the set* which is *associated with* $\mathfrak{B}$ *by the group pair* $(\mathfrak{U}, \mathfrak{V})$. Because of the symmetry of (20) we have

(7.2B) *If the group pair* $(\mathfrak{U}, \mathfrak{V})$ *is primitive, the relationship between* $\mathfrak{A}$ *and* $\mathfrak{B}$ *is reciprocal.*

Indeed, if we start from the mapping $A: u \rightarrow u^*$ of $\mathfrak{U}$, we see from (20) that the pair $(\mathfrak{U}, \mathfrak{V})$ admits the transformations of $\mathfrak{A}$, and that $\mathfrak{B}$ is the associated set.

3. Let $(\mathfrak{U}, \mathfrak{V})$ be again a group pair with 0 as the only l-annihilator. Every element u generates a homomorphic mapping $v \rightarrow uv$ of $\mathfrak{V}$ upon a subgroup of $\mathfrak{W}$ which is an operator-homomorphism with regard to the operators of Δ. All such operator-homomorphic mappings of $\mathfrak{V}$ upon a subgroup of $\mathfrak{W}$ form an additive group $\overline{\mathfrak{U}}$ which possesses the elements of Γ as l-operators. Then $\mathfrak{U}$ is (operator-) isomorphic with a subgroup of $\overline{\mathfrak{U}}$; we may consider $\mathfrak{U}$ itself as a subgroup of $\overline{\mathfrak{U}}$.

If $B: v \rightarrow v^*$ is a homomorphic mapping of $\mathfrak{V}$ upon $\mathfrak{V}$ or a subgroup of $\mathfrak{V}$, and if $\bar{u}$ is any element of $\overline{\mathfrak{U}}$, then $v \rightarrow \bar{u}v^*$ is an operator-homomorphic mapping of $\mathfrak{V}$ upon a subgroup of $\mathfrak{W}$ (with regard to the operators of Δ). It then is given by an element $\bar{u}^*$ of $\overline{\mathfrak{U}}$, and we have $\bar{u}^*v = \bar{u}v^*$. Hence

(7.3A) *If* $(\mathfrak{U}, \mathfrak{V})$ *is a group pair without nonzero l-annihilators, we can replace* $\mathfrak{U}$ *by a larger group* $\overline{\mathfrak{U}}$ *such that* $(\overline{\mathfrak{U}}, \mathfrak{V})$ *admits every set* $\mathfrak{B}$ *of homomorphic mappings of* $\mathfrak{V}$ *upon a subgroup of* $\mathfrak{V}$.

4. Let us restrict ourselves to the case that $\mathfrak{U}$ is a contragredient vector space and $\mathfrak{V}$ a cogredient vector space, the coordinates of the vectors taken from a fixed division ring K. We then take $\Gamma = \Delta = K$ in §7.1, and assume that $\mathfrak{W}$ is an m-dimensional cogredient vector space, and that l-multiplication of an element W with an element κ of K is performed by l-multiplying each component of W with κ[28]. We say in this case that $(\mathfrak{U}, \mathfrak{V})$ form a *group pair of rank* m. Assume that 0 is the only l-annihilator.

Let n be the number of dimensions of $\mathfrak{V}$. Since every element $\overline{U}$ of $\overline{\mathfrak{U}}$ corresponds to an operator-homomorphic mapping of $\mathfrak{V}$ upon a subgroup of $\mathfrak{W}$ (with regard to r-operators), it is given by a matrix of type (m, n) with coefficients in K. We may identify $\overline{U}$ with this matrix; the products $\kappa\overline{U}$ and $\overline{U}V$

([28]) We may then consider $\mathfrak{W}$ also as a contragredient vector space, if we consider only the addition in $\mathfrak{W}$ and the l-multiplication with elements of K. There will be no danger of a confusion, since we shall not perform linear transformations in $\mathfrak{W}$.

for κ in K, V in $\mathfrak{V}$ then have the ordinary significance (cf. §3). The number of dimensions of $\bar{\mathfrak{U}}$ is mn.

Every mapping B of $\mathfrak{V}$ of the kind considered in §7.3 is a linear transformation $V \to V^*$ and hence given by a matrix $(b_{\kappa\lambda})$ of degree n which we also denote by B setting $V^* = BV$. The associated mapping $\bar{A}: \bar{U} \to \bar{U}^*$ of $\mathfrak{U}$ is defined by $\bar{U}^* V = \bar{U} V^*$ or $\bar{U}^* V = \bar{U} B V$ which implies $\bar{U}^* = \bar{U} B$. This, of course, is a linear transformation $\bar{A}$ of $\bar{\mathfrak{U}}$ whose matrix we also denote by $\bar{A}$. We may consider $\bar{\mathfrak{U}}$ as a direct sum of m n-dimensional vector spaces $\mathfrak{T}_1, \cdots, \mathfrak{T}_m$ where in the matrices of $\mathfrak{T}_i$ only the coefficients in the ith row are different from 0. If we choose a basis $E_j^{(i)}$ of $\mathfrak{T}_i$ by taking the jth coefficients of the ith row equal to 1, and all the other coefficients equal to 0, we see that $\bar{A}$ transforms $E_j^{(i)}$ into $E_j^{(i)} B = \sum b_{j\lambda} E_\lambda^{(i)}$. This proves $\mathfrak{T}_i$ invariant under $\bar{A}$, the matrix of the induced transformation being B. Hence $\bar{A} = m \times B$. The set $\bar{\mathfrak{A}}$ associated with a set $\mathfrak{B}$ of transformations B by the pair $(\mathfrak{U}, \mathfrak{V})$ is then $\bar{\mathfrak{A}} = m \times \mathfrak{B}$.

If $\mathfrak{U}$ is a subgroup of $\bar{\mathfrak{U}}$, and the group pair $(\mathfrak{U}, \mathfrak{V})$ admits the transformations of $\mathfrak{B}$, then $\mathfrak{U}$ must be a subspace of $\bar{\mathfrak{U}}$ invariant under $\bar{\mathfrak{A}}$. The transformations of $\mathfrak{U}$ induced by $\bar{\mathfrak{A}}$ form a top constituent $\mathfrak{A}$ of $\bar{\mathfrak{A}}$, and this $\mathfrak{A}$ is the set associated with $\mathfrak{B}$ by the group pair $(\mathfrak{U}, \mathfrak{V})$. Hence (cf. §4.3)

(7.4A) *Let $\mathfrak{U}$ be a contragredient vector space and $\mathfrak{V}$ a cogredient vector space both forming a group pair of rank m. If 0 is the only l-annihilator, and $(\mathfrak{U}, \mathfrak{V})$ admits the set $\mathfrak{B}$ of homomorphic mappings of $\mathfrak{V}$ upon $\mathfrak{V}$ or a subgroup of $\mathfrak{V}$, then the associated set $\mathfrak{A}$ is a top constituent of $m \times \mathfrak{B}$.*

In the same manner, we prove

(7.4B) *If 0 is the only r-annihilator in $(\mathfrak{U}, \mathfrak{V})$, and $(\mathfrak{U}, \mathfrak{V})$ admits the set $\mathfrak{A}$ of homomorphic mappings of $\mathfrak{U}$ upon a subgroup of $\mathfrak{U}$, then the associated set of transformations of $\mathfrak{B}$ is an end constituent of $m \times \mathfrak{A}$.*

That we here obtain an end constituent instead of a top constituent as in (7.4A) is due to the fact that $\mathfrak{V}$ is a cogredient vector space. The transformations induced in an invariant subspace are end constituents (cf. §4.3).

5. Let us apply the preceding considerations to sets $\mathfrak{B}$ of matrices of degree n with coefficients in the division ring K. Let $m > 0$ be a given integer. We say that a set $\mathfrak{U}$ of matrices of type (m, n) with coefficients in K is a $(K, \mathfrak{B})$-*double module*, if $\mathfrak{U}$ contains the matrices $U_1 + U_2$, κU, UB for any U, U_1, U_2 in $\mathfrak{U}$, any κ in K, and any B in $\mathfrak{B}$. We then choose an l-basis $U_1, U_2, \cdots, U_k$ of $\mathfrak{U}$. Since any product $U_\kappa B$ lies in $\mathfrak{U}$ again, we have formulae

$$U_\kappa B = \sum_{\lambda=1}^{k} a_{\kappa\lambda} U_\lambda, \qquad \kappa = 1, 2, \cdots, k, \tag{21}$$

with coefficients $a_{\kappa\lambda}$ in K. We say that the set $\mathfrak{A}$ of all the matrices $A = (a_{\kappa\lambda})$ is the set *associated with $\mathfrak{B}$ by the double module* $\mathfrak{U}$. The degree k of $\mathfrak{A}$ is the

l-rank of $\mathfrak{U}$. If $\mathfrak{B}$ is closed under addition or multiplication, the set $\mathfrak{A}$ is homomorphic with $\mathfrak{B}$ with regard to this operation([29]). If the l-basis U_κ is replaced by another l-basis, $\mathfrak{A}$ is replaced by a similar set.

If $\mathfrak{V}$ is the n-dimensional cogredient vector space in which the transformations of $\mathfrak{B}$ take place, then $(\mathfrak{U}, \mathfrak{V})$ form a group pair, the product UV of a matrix U of $\mathfrak{U}$ and a vector $\mathfrak{V}$ being defined in the ordinary manner. This group pair $(\mathfrak{U}, \mathfrak{V})$ is of rank m, and 0 is the only l-annihilator.

Further, $(\mathfrak{U}, \mathfrak{V})$ admits the transformations $\mathfrak{B}$ of $\mathfrak{V}$, and $\mathfrak{A}$ is the associated set in the sense of §7.2, since the transformation $U_\kappa \rightarrow U_\kappa B$ in the contragedient vectors space with the basis $U_1, U_2, \cdots, U_k$ has the matrix A according to (21). From (7.4A) there follows

(7.5A) *If $\mathfrak{U}$ is a $(K, \mathfrak{B})$-double module, consisting of matrices of type (m, n), then $\mathfrak{U}$ associates the set of matrices $\mathfrak{B}$ of degree n with a set $\mathfrak{A}$ which is a top constituent of $m\times\mathfrak{B}$.*

The r-annihilators of $(\mathfrak{U}, \mathfrak{V})$ will form a subspace $\mathfrak{V}_0$ of $\mathfrak{V}$ which is invariant under $\mathfrak{B}$ since $\mathfrak{U}\cdot BV_0 \subseteq \mathfrak{U}V_0 = (0)$ for V_0 in $\mathfrak{V}_0$, B in $\mathfrak{B}$. Let $\mathfrak{B}_0$ be the set of transformations of $\mathfrak{V}/\mathfrak{V}_0$ induced by $\mathfrak{B}$; then $\mathfrak{B}_0$ is a top constituent of $\mathfrak{B}$ according to §4.3. We may consider $(\mathfrak{U}, \mathfrak{V}/\mathfrak{V}_0)$ as a primitive group pair consisting of a contragredient vector space $\mathfrak{U}$ and a cogredient vector space $\mathfrak{V}/\mathfrak{V}_0$. The rank of this group pair still is m. If $B: V \rightarrow V^*$ is a transformation of $\mathfrak{B}$, and $A: U \rightarrow U^*$ the corresponding transformation of $\mathfrak{A}$, then we have $U^*V = UV^*$. The corresponding equation holds, when we replace V and V^* by their residue class modulo $\mathfrak{V}_0$. Then $V^* \equiv B_0 V \pmod{\mathfrak{V}_0}$ where B_0 is the matrix of $\mathfrak{B}_0$ corresponding to B in $\mathfrak{B}$. Consequently, the group pair $(\mathfrak{U}, \mathfrak{V}/\mathfrak{V}_0)$ associates the set of transformations $\mathfrak{B}_0$ of $\mathfrak{V}/\mathfrak{V}_0$ with the set $\mathfrak{A}$ of transformations of $\mathfrak{U}$ and vice versa (cf. (7.2B)).

Then from (7.4B) we obtain

(7.5B) *In* (7.5A) *let $\mathfrak{V}_0$ be the set of all n-dimensional vectors V_0 for which $UV_0 = 0$ for every U in $\mathfrak{U}$. Then $\mathfrak{V}_0$ is invariant under $\mathfrak{B}$. If $\mathfrak{B}_0$ is the top constituent of $\mathfrak{B}$, consisting of the transformations of $\mathfrak{V}/\mathfrak{V}_0$ induced by $\mathfrak{B}$, then $\mathfrak{B}_0$ is an end constituent of $m\times\mathfrak{A}$.*

6. We can now apply (6.6A), (6.6B), and (6.6C) and obtain

(7.6A) *In the notation of* (7.5A) *and* (7.5B) *$\mathfrak{A}$ and $\mathfrak{B}_0$ have the same number of Loewy constituents: $L(\mathfrak{A}) = L(\mathfrak{B}_0)$. Every irreducible constituent of $\mathfrak{L}_i(\mathfrak{B}_0)$ appears in $\mathfrak{L}_i(\mathfrak{A})$, and every irreducible constituent of $\mathfrak{L}_i(\mathfrak{A})$ appears in some $\mathfrak{L}_{i+j}(\mathfrak{B}_0)$ with $j \geqq 0$. Every irreducible constituent $\tilde{\mathfrak{L}}_i(\mathfrak{A})$ appears in $\tilde{\mathfrak{L}}_i(\mathfrak{B}_0)$, and every irreducible constituent of $\tilde{\mathfrak{L}}_i(\mathfrak{B}_0)$ appears in some $\tilde{\mathfrak{L}}_{i+j}(\mathfrak{A})$ with $j \geqq 0$.*

We have the corollary

([29]) In the notation of E. Noether [20], $\mathfrak{U}$ is a representation module for the representation $\mathfrak{A}$ of $\mathfrak{B}$.

(7.6B) *The sets $\mathfrak{A}$ and $\mathfrak{B}_0$ have the same irreducible constituents though not necessarily with the same multiplicities.*

It is also possible to make some statements concerning the multiplicities, e.g.,

(7.6C) *If an irreducible constituent $\mathfrak{F}$ appears h times in $L_i(\mathfrak{B}_0)$, it appears at least h/m times in $L_i(\mathfrak{A})$. (Similarly in the other cases.)*

7. As an application, we prove the following theorem:

(7.7A) *Let $\mathfrak{B}$ be a set of matrices which has no constituents (0), and let $\mathfrak{U}$ be a $(K, \mathfrak{B})$ double module consisting of matrices of type (m, n). The necessary and sufficient condition that a matrix Z of type (m, n) belongs to $\mathfrak{U}$ is that ZB belongs to $\mathfrak{U}$ for every B in $\mathfrak{B}$.*

Proof. Let $U_1, U_2, \cdots, U_k$ be an l-basis of $\mathfrak{U}$. If Z does not belong to $\mathfrak{U}$, then $U_1, U_2, \cdots, U_k, Z$ will be an l-basis of a $(K, \mathfrak{B})$-double module $\mathfrak{U}^*$. The set associated with $\mathfrak{B}$ by $\mathfrak{U}^*$ has the form

$$\mathfrak{A}^* = \begin{pmatrix} \mathfrak{A} & \\ * & 0 \end{pmatrix},$$

where $\mathfrak{A}$ is the associated set with $\mathfrak{B}$ by $\mathfrak{U}$. According to (7.6A) every irreducible constituent of $\mathfrak{A}^*$ must appear in $\mathfrak{B}$ whereas 0 is no constituent of $\mathfrak{B}$. Hence Z must belong to $\mathfrak{U}$.

8. Finally, we give some formulae showing the relationship between $\mathfrak{A}$ and $\mathfrak{B}$ in a more formal manner.

(7.8A) *Let $\mathfrak{A}$ and $\mathfrak{B}$ be two related sets of matrices of degrees k and n respectively, let m be a positive integer, and $h_{\mu\nu}^{(\kappa)}$ a set of kmn elements of K $(\kappa=1, 2, \cdots, k;\ \mu=1, 2, \cdots, m;\ \nu=1, 2, \cdots, n)$. We form three sets of matrices, U_κ of type (m, n), T_ν of type (k, m) and P_μ of type (k, n):*

$$U_\kappa = (h_{\alpha\beta}^{(\kappa)}); \qquad T_\nu = (h_{\beta\nu}^{(\alpha)}); \qquad P_\mu = (h_{\mu\beta}^{(\alpha)}), \tag{22}$$

where α is the row index and β the column index. The three sets of relations (for corresponding $A = (a_{\alpha\beta})$ and $B = (b_{\alpha\beta})$)

$$U_\alpha B = \sum a_{\alpha\beta} U_\beta, \tag{23a}$$

$$AT_\nu = \sum T_\lambda b_{\lambda\nu}, \tag{23b}$$

$$\mathfrak{A}P_\mu = P_\mu \mathfrak{B} \tag{23c}$$

are equivalent.

Proof. All three relations are equivalent to

$$\sum_\lambda h_{\mu\lambda}^{(\alpha)} b_{\lambda\nu} = \sum_\beta a_{\alpha\beta} h_{\mu\nu}^{(\beta)}.$$

The equation (23a) is identical with (21). The equation (23b) shows that all matrices of the form $\sum T_\nu c_\nu$, c_ν in K, form what we may call an $(\mathfrak{A}, K)$ double module $\mathfrak{T}$. If the T_ν are r-independent, this $\mathfrak{T}$ associates $\mathfrak{A}$ with $\mathfrak{B}$, and this again expresses the reciprocity between $\mathfrak{B}$ and $\mathfrak{A}$.

8. The regular representation

1. We now consider a set $\mathfrak{G}$ of square matrices which forms a *semi-group*, i.e., which contains the product of any two of its matrices. Let $U_1, U_2, \cdots, U_k$ be an l-basis of $\mathfrak{G}$. The linear combinations $\sum c_\kappa U_\kappa$ with arbitrary coefficients in K form a $(K, \mathfrak{G})$-double module which we call the *enveloping module* $\mathfrak{M}(\mathfrak{G})$ of $\mathfrak{G}$. For G in $\mathfrak{G}$, we have the formulae

$$U_\kappa G = \sum_\lambda r_{\kappa\lambda} U_\lambda, \qquad r_{\kappa\lambda} \text{ in } K, \tag{24}$$

and the matrices $R = (r_{\kappa\lambda})$ form the associated set $\mathfrak{R}$. The mapping $G \rightarrow R$ is a homomorphism with regard to multiplication. In other words, $\mathfrak{R}$ is a representation of $\mathfrak{G}$, known as the *regular representation*(30) of $\mathfrak{G}$. If the l-basis U_κ is replaced by another l-basis of $\mathfrak{M}(\mathfrak{G})$, then $\mathfrak{R}$ is replaced by a similar set(31). The degree of the regular representation is equal to the l-rank of $\mathfrak{G}$.

2. Let $\mathfrak{B}$ be the space in which the transformations of $\mathfrak{G}$ take place. We shall apply (7.5A) and (7.5B) (for $\mathfrak{U} = \mathfrak{M}(\mathfrak{G})$). Here $\mathfrak{B}_0$ consists of those vectors V for which $\mathfrak{M}(\mathfrak{G})V = 0$. This condition is equivalent with $\mathfrak{G}V = 0$, and hence $\mathfrak{G}$ induces the transformation 0 in $\mathfrak{B}_0$. It follows that in a suitable coordinate system

$$\mathfrak{G} \sim \begin{pmatrix} \mathfrak{G}_0 & \\ * & 0 \end{pmatrix}, \tag{25}$$

where the constituent 0 at the bottom is of degree $n_0 \geqq 0$(32). It is not possible to find a similar set with a bottom constituent 0 of higher degree. From the theorems in §7.5 and §7.6 we derive:

(8.2A) *Let $\mathfrak{G}$ be a semigroup of matrices of degree n. We split $\mathfrak{G}$ into a constituent $\mathfrak{G}_0$ and a bottom constituent 0 of highest possible degree, (25). The regular representation $\mathfrak{R}$ of $\mathfrak{G}$ is a top constituent of $n \times \mathfrak{G}_0$, and $\mathfrak{G}_0$ is an end constituent of $n \times \mathfrak{R}$.*

(8.2B) *We have $L(\mathfrak{R}) = L(\mathfrak{G}_0)$. Every irreducible constituent of $L_i(\mathfrak{G}_0)$ appears in $\mathfrak{L}_i(\mathfrak{R})$, and every irreducible constituent of $\mathfrak{L}_i(\mathfrak{R})$ appears in some $\mathfrak{L}_{i+j}(\mathfrak{G}_0)$*

(30) For properties of the regular representation, cf. Frobenius [9], MacDuffee [16], Brauer and Nesbitt [4], Nesbitt [19], Nakayama [18].

(31) It should be noticed that in the case of a non-commutative K, the module $\mathfrak{M}(\mathfrak{G})$ is, in general, not a ring. Further, similar semi-groups $\mathfrak{G}$ and $\mathfrak{G}_1$ may have different l-ranks and different regular representations.

(32) If $n_0 = 0$, then the constituent 0 in (25) is missing.

with $j \geqq 0$. Every irreducible constituent of $\tilde{\mathfrak{L}}_i(\mathfrak{R})$ appears in $\tilde{\mathfrak{L}}_i(\mathfrak{G}_0)$, and every irreducible constituent of $\tilde{\mathfrak{L}}_i(\mathfrak{G}_0)$ appears in some $\tilde{\mathfrak{L}}_{i+j}(\mathfrak{R})$ with $j \geqq 0$.

These results lead to the following corollaries:

(8.2C) *We have either $L(\mathfrak{R}) = L(\mathfrak{G})$, or $L(\mathfrak{R}) = L(\mathfrak{G}) - 1$. If $\mathfrak{L}_1(\mathfrak{G})$ does not contain a constituent 0, we have the first case.*

(8.2D) *The sets $\mathfrak{G}$ and $\mathfrak{R}$ have the same irreducible constituents, except perhaps constituents 0 which may appear in $\mathfrak{G}$ without appearing in $\mathfrak{R}$.*

(8.2E) *If $\mathfrak{G}$ is completely reducible, so is $\mathfrak{R}$.*

For $L(\mathfrak{G}) = 1$ implies $L(\mathfrak{R}) = 1$ by (8.2C), and this is equivalent to the complete reducibility of $\mathfrak{R}$.

In certain cases, $\mathfrak{G}_0$ can be replaced by $\mathfrak{G}$. We can prove

(8.2F) *If $\mathfrak{M}(\mathfrak{G})$ contains a matrix $J \neq 0$ such that $JG = G$ for every G in $\mathfrak{G}$, then $\mathfrak{G}$ splits completely into $\mathfrak{G}_0$ and a constituent 0, and we have $L(\mathfrak{R}) = L(\mathfrak{G}) = L(\mathfrak{G}_0)$, $\mathfrak{L}_i(\mathfrak{G}) = \mathfrak{L}_i(\mathfrak{G}_0)$ for every $i \geqq 2$. The assumption is satisfied, in particular, when $\mathfrak{G}$ has a l-unit J.*

Proof. Assume that $Q^{-1}\mathfrak{G}Q$ splits in the form (25). The last n_0 columns in all the matrices of $Q^{-1}\mathfrak{G}Q$ vanish. The same then is true for $Q \cdot Q^{-1}\mathfrak{G}Q = \mathfrak{G}Q$, hence for $\mathfrak{M}(\mathfrak{G})Q$, and for $Q^{-1}\mathfrak{M}(\mathfrak{G})Q$. We may set

$$Q^{-1}\mathfrak{G}Q = \begin{pmatrix} \mathfrak{G}_0 & \\ \mathfrak{C} & 0 \end{pmatrix}, \qquad Q^{-1}JQ = \begin{pmatrix} X & 0 \\ Y & 0 \end{pmatrix}$$

since J belongs to $\mathfrak{M}(\mathfrak{G})$. From $JG = G$, we obtain $Y\mathfrak{G}_0 = \mathfrak{C}$ or $\mathfrak{C} = Y\mathfrak{G}_0 - 0Y$. This shows that after an elementary similarity transformation, we may replace $\mathfrak{C}$ by 0. This shows the first part of (8.2F); the other statements follow from it.

From (7.7A), we obtain at once

(8.2G) *Let $\mathfrak{G}$ be a semigroup of matrices of degree n which has no constituent 0. A necessary and sufficient condition that a matrix Z of degree n belongs to $\mathfrak{M}(\mathfrak{G})$ is that ZG belongs to $\mathfrak{M}(\mathfrak{G})$ for every G in $\mathfrak{G}$. In particular, the unit matrix I belongs to $\mathfrak{M}(\mathfrak{G})$.*

3. In certain cases, the theorem (8.2B) can be improved. We prove:

(8.3A) *Assume that the semigroup $\mathfrak{G}$ itself appears in its lower Loewy normal form, and that no constituent 0 appears in $\mathfrak{G}$. Every irreducible constituent of $\mathfrak{L}_i(\mathfrak{G})$ is also a constituent of $\mathfrak{L}_1(\mathfrak{R})$, $L_2(\mathfrak{R})$, $\cdots$, $\mathfrak{L}_i(\mathfrak{R})$.*

Proof. Assume that the semigroup $\mathfrak{G}$ itself splits into several constituents, one of which is $\mathfrak{H}$. Denote by $W_1, W_2, \cdots, W_k$ the matrices of $\mathfrak{M}(\mathfrak{H})$ which correspond to an l-basis $U_1, U_2, \cdots, U_k$ of $\mathfrak{M}(\mathfrak{G})$. Obviously, we can choose

$U_1, U_2, \cdots, U_k$ such that $W_1 = \cdots = W_j = 0$ and $W_{j+1}, \cdots, W_k$ form an l-basis of $\mathfrak{M}(\mathfrak{H})$. If G in $\mathfrak{G}$ corresponds to H in $\mathfrak{H}$, then (24) implies

$$W_\kappa H = \sum r_{\kappa\lambda} W_\lambda$$

and we easily see that $\mathfrak{R}$ splits into a top constituent of degree j and the regular representation $\mathfrak{R}^*$ of $\mathfrak{H}$ as end constituent. From (6.6A) and (8.2B) it follows that every irreducible constituent of $\mathfrak{L}_\nu(\mathfrak{H})$ appears in $\mathfrak{L}_\nu(\mathfrak{R})$; (0) is not a constituent of $\mathfrak{H}$.

We now choose $\mathfrak{H}$ as the constituent of $\mathfrak{G}$ which contains the Loewy constituents $\mathfrak{L}_L(\mathfrak{G}), \cdots, \mathfrak{L}_\beta(\mathfrak{G})$. Then $\mathfrak{L}_\nu(\mathfrak{H}) = \mathfrak{L}_{\beta+\nu-1}(\mathfrak{G})$, and for $i = \nu+\beta-1$, $1 \leqq \nu \leqq i$, we obtain the statement of (8.3A).

If the underlying division ring is a field, there is no restriction in the assumption that $\mathfrak{G}$ itself is in its lower Loewy normal form, since similar semigroups here have the same regular representation.

4. A discussion, analogous to that in §8.1, is possible with regard to an r-basis $\tilde{U}_1, \cdots, \tilde{U}_l$ of $\mathfrak{G}$. Here we set

$$G\tilde{U}_\lambda = \sum_\kappa \tilde{U}_\kappa s_{\kappa\lambda}, \qquad s_{\kappa\lambda} \text{ in } K, \tag{26}$$

and $G \rightarrow S = (s_{\kappa\lambda})$ defines the *second regular representation* of $\mathfrak{G}$. Going over to transposed matrices (cf. (3.5)) in (26), we obtain

(8.4A) *The second regular representation $\mathfrak{S}$ of a semigroup $\mathfrak{G}$ is the transpose of the first regular representation of the transposed set $\mathfrak{G}'$.*

This remark allows us to restrict ourselves to the consideration of the first regular representation.

9. Irreducible semigroups

1. We now consider irreducible semigroups $\mathfrak{G} \neq (0)$ consisting of square matrices of degree n with coefficients in the division ring K. Since the degree of the regular representation equals the l-rank of $\mathfrak{G}$, we obtain from (8.2B):

(9.1A) *If $\mathfrak{G}$ is an irreducible semigroup of degree n and l-rank k, then the regular representation $\mathfrak{R}$ of $\mathfrak{G}$ is similar to $(k/n) \times \mathfrak{G}$. In particular, the l-rank is a multiple of the degree.*

We wish to characterize the number k/n by means of the commuting ring $\mathfrak{C}(\mathfrak{G})$ of $\mathfrak{G}$. Denoting the row $(0, \cdots, 0, 1, 0, \cdots, 0)$ with the ith component 1 by E_i, we see that E_iC is the ith row of the matrix C. We determine the largest number h of indices $\mu_1, \mu_2, \cdots, \mu_h$, with $1 \leqq \mu_i \leqq n$, such that conditions

$$\sum_\mu E_\mu C_\mu = 0, \qquad C_\mu \text{ in } \mathfrak{C}(\mathfrak{G}),\ \mu \text{ ranging over } \mu_1, \cdots, \mu_h, \tag{27}$$

imply $C_{\mu_i} = 0$ for all μ_i. Since all the $C_\mu \neq 0$ in $\mathfrak{C}(\mathfrak{G})$ are nonsingular (cf.

(6.2A)), we have $h \geqq 1$. The μ_i are all distinct, since if for example $\mu_1 = \mu_2$, we could set $C_{\mu_1} = -C_{\mu_2} \neq 0$, and all the later $C_\mu = 0$ in (27). We denote this number h as the *h-number of* $\mathfrak{C}(\mathfrak{G})$, and state

(9.1B) *The quotient k/n in* (9.1A) *is equal to the h-number of* $\mathfrak{C}(\mathfrak{G})$.

Proof. Assume first that $h < n$. For any fixed $i = 1, 2, \cdots, n$, we can find matrices $C_{\mu i}$ $(\mu = \mu_1, \cdots, \mu_h)$ and C_i in $\mathfrak{C}(\mathfrak{G})$ such that

$$\sum_\mu E_\mu C_{\mu i} + E_i C_i = 0 \tag{28}$$

and not all $C_{\mu i}$, C_i vanish. Then $C_i \neq 0$, because otherwise (28) would be identical with (27) for $C_{\mu i} = C_\mu$, and all these matrices would also vanish. Because of (6.2A), C_i is nonsingular, and if we r-multiply (28) by its reciprocal, we see that we may assume $C_i = I$. We then multiply (28) by an arbitrary element G of $\mathfrak{G}$, and obtain

$$0 = \sum_\mu E_\mu C_{\mu i} G + E_i G = \sum_\mu E_\mu G C_{\mu i} + E_i G. \tag{29}$$

Denote by $t_1, t_2, \cdots, t_{hn}$ the hn coefficients appearing in the rows $\mu_1, \mu_2, \cdots, \mu_h$ of G. Since $E_i G$ is the ith row of G, and $E_\mu G$ the μth row of G, we see from (29) that every fixed coefficient of G, say in the ith row and jth column, is a linear function $\sum t_\rho \gamma_\rho$, where the γ_ρ are elements of K which are independent of G (but dependent on i, j). Then G has the form $G = \sum t_\rho Q_\rho$, where the Q_ρ are fixed matrices, and this shows that the l-rank k of $\mathfrak{G}$ is not larger than hn. This is also true, if $h = n$, since certainly $k \leqq n^2$. Thus we always have $k/n \leqq h$.

On the other hand, we may choose an l-basis $U_1, U_2, \cdots, U_k$ of $\mathfrak{M}(\mathfrak{G})$, such that the regular representation $\mathfrak{R}$ with regard to this basis has the form (cf. (9.1A))

$$\mathfrak{R} = j \times \mathfrak{G}, \qquad j = k/n. \tag{30}$$

We now apply (7.8A) to $\mathfrak{A} = \mathfrak{R}$ and $\mathfrak{B} = \mathfrak{G}$, using for U_κ the notation of the first formula (22) and defining P_μ by the last formula (22); we have here $m = n$. For the n matrices P_μ which intertwine $\mathfrak{R}$ and $\mathfrak{G}$ $(\mu = 1, 2, \cdots, n)$, according to (22), we have

$$E_\nu P_\mu = E_\mu U_\nu = (h_{\mu 1}^{(\nu)}, h_{\mu 2}^{(\nu)}, \cdots, h_{\mu n}^{(\nu)}). \tag{31}$$

We break up each matrix P_μ according to the scheme $(n, n, \cdots, n | n)$,

$$P_\mu = \begin{pmatrix} Q_{\mu 1} \\ \cdot \\ \cdot \\ \cdot \\ Q_{\mu j} \end{pmatrix}. \tag{32}$$

Because of (30), each $Q_{\mu\rho}$ intertwines $\mathfrak{G}$ with $\mathfrak{G}$; i.e., $Q_{\mu\rho}$ belongs to $\mathfrak{C}(\mathfrak{G})$.

Choose any $j+1$ values μ from $1, 2, \cdots, n$, and consider the j linear equations

$$\sum Q_{\mu 1} X_\mu = 0, \cdots, \sum Q_{\mu j} X_\mu = 0.$$

Since the coefficients lie in the division ring $\mathfrak{C}(\mathfrak{G})$, and we have more unknowns X_i than equations, there is a non-trivial solution X_i in $\mathfrak{C}(\mathfrak{G})$ (cf. (3.4)). Then $\sum P_\mu X_\mu = 0$. On l-multiplying by E_ν and using (31), we obtain

$$\sum_\mu E_\mu U_\nu X_\mu = 0. \tag{33}$$

We now determine $z_1, \cdots, z_n$ in K such that $\sum z_\nu U_\nu = I$. This is possible (cf. (8.2G)). Since $z_\nu E_\mu = E_\mu z_\nu$, l-multiplication of (33) with z_ν and addition over ν yields $\sum E_\mu X_\mu = 0$. Since the X_μ are elements of $\mathfrak{C}(\mathfrak{G})$ which do not all vanish, this is a relation (27). For any $j+1$ indices μ, we have a non-trivial relation of this kind. Hence $j+1 > h$, i.e., $j \geqq h$. Because of (30), we have $k/n \geqq h$. Since we also showed $k/n \leqq h$, the statement is proved.

In the notation of the first part of this proof, it now follows that the matrices Q_ρ are l-independent and belong to $\mathfrak{M}(\mathfrak{G})$, since otherwise $\mathfrak{M}(\mathfrak{G})$ would have an l-rank smaller than hn. Further, $t_1, \cdots, t_{hn}$ are the coefficients in the rows $\mu_1, \cdots, \mu_h$ of $\sum t_\rho Q_\rho$. Hence

(9.1C) *In the notation of* (9.1B), *the coefficients in h suitable rows $\mu_1, \cdots, \mu_h$ of a matrix M of $\mathfrak{M}(\mathfrak{G})$ can be assigned as arbitrary elements of K, and then M is determined uniquely. We can choose the indices μ as in* (27).

2. Let v be the r-rank of $\mathfrak{C}(\mathfrak{G})$. There exist at most jv matrices (32) which are r-independent, since $Q_{\mu\rho}$ lies in $\mathfrak{C}(\mathfrak{G})$, where $j = k/n = h$. If we now choose more than hv distinct indices μ from $1, 2, \cdots, n$ (assuming that $n > hv$), then the matrices P_μ are r-dependent and we have equations $\sum P_\mu x_\mu = 0$ (x_μ in K, not all of them 0). We proceed as in the second part of the proof of (9.1B). On l-multiplying with E_ν and using (31), we find $\sum E_\mu U_\nu k_\mu = 0$ (summed over μ). Again, l-multiplying by the same z_ν as above and adding, we find $\sum E_\mu x_\mu = 0$. But this implies $x_\mu = 0$, which gives a contradiction. Hence $n \leqq hv$, which gives

(9.2A) *Let $\mathfrak{G}$ be an irreducible semigroup of degree n. If $\mathfrak{G}$ has the l-rank k, and $\mathfrak{C}(\mathfrak{G})$ has the r-rank v, then $n^2 \leqq kv$.*

This can be considered as a generalization of Burnside's theorem (cf. §9.4).

3. Consider a similarity transformation applied to the irreducible semigroup $\mathfrak{G}$. The same transformation, then, is to be applied to $\mathfrak{C}(\mathfrak{G})$. According to (6.4B), the set $\mathfrak{C}(\mathfrak{G})$ has only one irreducible constituent $\mathfrak{W}$, and after the similarity transformation, we may assume that

$$\mathfrak{C}(\mathfrak{G}) = s \times \mathfrak{W} \tag{34}$$

where $n/s = t$ is the degree of $\mathfrak{W}$.

We set $\mathfrak{C}(\mathfrak{W})=\mathfrak{T}$. Since $\mathfrak{W}$ is irreducible, $\mathfrak{T}$ is a division ring. From (6.2B)

$$\mathfrak{C}\mathfrak{C}(\mathfrak{G}) = \mathfrak{C}(s \times \mathfrak{W}) = [\mathfrak{C}(\mathfrak{W})]_s = [\mathfrak{T}]_s;$$

and since $\mathfrak{G}\subseteq\mathfrak{C}\mathfrak{C}(\mathfrak{G})$, we have

$$\mathfrak{G} \subseteq [\mathfrak{T}]_s.$$

The irreducibility of $\mathfrak{G}$ implies the irreducibility of $\mathfrak{T}$, from the first part of theorem (6.3A). Obviously, $\mathfrak{C}(\mathfrak{T})\supseteq\mathfrak{W}$. If we had $\mathfrak{C}(\mathfrak{T})\supset\mathfrak{W}$, then, according to (6.2B) we would have $\mathfrak{C}(\mathfrak{G})\supseteq\mathfrak{C}([\mathfrak{T}]_s)=s\times\mathfrak{C}(\mathfrak{T})\supset s\times\mathfrak{W}=\mathfrak{C}(\mathfrak{G})$, which is impossible. Hence $\mathfrak{W}$ and $\mathfrak{T}$ both are irreducible division rings consisting of matrices of degree t, and each is the commuting set of the other.

We now apply theorem (9.1B) to $\mathfrak{T}$ instead of $\mathfrak{G}$. If h_0 is the h-number of $\mathfrak{C}(\mathfrak{T})=\mathfrak{W}$, and z the l-rank of $\mathfrak{T}$, then $h_0=z/t$. But (34) shows that the h-number of $\mathfrak{C}(\mathfrak{G})$ is $h=sh_0$, and hence

$$k/n = h = sh_0 = sz/t \tag{35}$$

which implies $k=s^2z$ since $n=st$. Consequently, $\mathfrak{G}$ and $[\mathfrak{T}]_s$ have the same l-rank, and therefore $\mathfrak{M}(\mathfrak{G})=\mathfrak{M}([\mathfrak{T}]_s)$. Thus we have

(9.3A) *Any irreducible semigroup $\mathfrak{G}$ of degree n is, after a similarity transformation, contained in a set $[\mathfrak{T}]_s$ where $\mathfrak{T}$ is an irreducible set of matrices of degree $n/s=t$ forming a division ring, and $\mathfrak{G}$ and $[\mathfrak{T}]_s$ have the same l-rank and hence the same enveloping module, $\mathfrak{M}(\mathfrak{G})=\mathfrak{M}([\mathfrak{T}]_s)$. Further, $\mathfrak{W}=\mathfrak{C}(\mathfrak{T})$ is the only irreducible constituent of $\mathfrak{C}(\mathfrak{G})$ and its multiplicity is s, i.e., $\mathfrak{C}(\mathfrak{G})=s\times\mathfrak{W}$. Conversely, $\mathfrak{T}=\mathfrak{C}(\mathfrak{W})$.*

Let v be the l-rank of $\mathfrak{C}(\mathfrak{G})$ which by (34) is also the l-rank of $\mathfrak{W}$, and let z be the l-rank of $\mathfrak{T}$. From (35), we obtain

$$\frac{kv}{n^2} = \frac{sz}{t}\,\frac{v}{n} = \frac{z}{t}\,\frac{v}{t}.$$

Both fractions on the right side are integers; they give the multiplicity of $\mathfrak{T}$ and of $\mathfrak{W}$ in their regular representations. The same is true if we take for v the r-rank of $\mathfrak{C}(\mathfrak{G})$. Then v/t is the multiplicity of $\mathfrak{W}$ in its second regular representation. Hence we have

(9.3B) *If in* (9.3A) *the set $\mathfrak{T}$ has the l-rank z, if $\mathfrak{G}$ has the l-rank k, and $\mathfrak{W}$ the l-rank v (or the r-rank v), then $kv/n^2=(z/t)(v/t)$ where z/t and v/t are integers.*

This gives, of course, the inequality of (9.2A); but it is not sufficient for a proof of (9.2A) in the general case, since we applied here a similarity transformation which may have changed the original ranks.

4. If the underlying division ring K is a field[33], then l-rank and r-rank

[33] For this case, compare, for instance, Weyl [31].

always coincide. Further $\mathfrak{M}(\mathfrak{T})=\mathfrak{T}$, since every linear combination of elements of $\mathfrak{T}$ commutes with every element $\mathfrak{W}$. Similarly, $\mathfrak{M}(\mathfrak{W})=\mathfrak{W}$.

If $\mathfrak{G}$ is an irreducible algebra of matrices, then $\mathfrak{M}(\mathfrak{G})=\mathfrak{G}$, and (9.3A) shows that $\mathfrak{G}\sim[\mathfrak{T}]_s$ where $\mathfrak{T}$ itself is an irreducible division algebra over K. This is Wedderburn's theorem.

For an irreducible division algebra $\mathfrak{G}$ of matrices, the number $h=k/n$ must be equal to 1; as follows for instance from (9.1C) since here it is certainly impossible to choose the coefficients in two rows arbitrarily. For such a $\mathfrak{G}$ the rank k and the degree n are equal.

If we apply this to $\mathfrak{T}$ and $\mathfrak{W}$ in (9.3B), we have $z=t$ and $v=t$ and hence

$$kv = n^2, \tag{36}$$

where n is the degree of the irreducible semigroup $\mathfrak{G}$, k is the rank of $\mathfrak{G}$, and v the rank of $\mathfrak{C}(\mathfrak{G})$. This is the generalized Burnside theorem. We obtain the original theorem when we assume that the field K is algebraically closed, and therefore $v=1$, i.e., $k=n^2$. This can also be derived from (9.2A),

We also obtain

(9.4A) *If K is a field, and $\mathfrak{G}$ an irreducible algebra of matrices, we have* $\mathfrak{C}(\mathfrak{C}(\mathfrak{G}))=\mathfrak{G}$.

Proof. We have $\mathfrak{M}(\mathfrak{G})=\mathfrak{G}$, and, because of the commutativity of K, this is not affected by a similarity transformation. We may assume $\mathfrak{G}$ in the form $\mathfrak{G}=[\mathfrak{T}]_s$. Further, $\mathfrak{M}(\mathfrak{T})=\mathfrak{T}$. Then (9.3A) in connection with (6.2B) gives $\mathfrak{C}(\mathfrak{C}(\mathfrak{G}))=\mathfrak{C}(s\times\mathfrak{W})=[\mathfrak{C}(\mathfrak{W})]_s=[\mathfrak{T}]_s=\mathfrak{G}$. The same equation $\mathfrak{C}\mathfrak{C}(\mathfrak{G})=\mathfrak{G}$ must have been true then, before $\mathfrak{G}$ was subjected to the similarity transformation mentioned in (9.3A).

10. On the representation of sets of matrices as direct sums. The radical

1. We say that a set $\mathfrak{Q}$ of square matrices of degree n is the *sum of two subsets* $\mathfrak{A}$ and $\mathfrak{B}$, if $\mathfrak{Q}$ consists of all the matrices $A+B$ with A in $\mathfrak{A}$, B in $\mathfrak{B}$. We write $\mathfrak{Q}=\mathfrak{A}\oplus\mathfrak{B}$, if, besides, we have $\mathfrak{A}\mathfrak{B}=0$ and $\mathfrak{B}\mathfrak{A}=0$ (i.e., $AB=0$ and $BA=0$ for any A in $\mathfrak{A}$ and any B in $\mathfrak{B}$)([34]). We first prove

(10.1A) *If the semigroup $\mathfrak{G}$ breaks up completely into m distinct (i.e., non-similar) irreducible constituents*

$$\mathfrak{G} = \begin{vmatrix} \mathfrak{F}_1 & & & \\ & \mathfrak{F}_2 & & \\ & & \cdot & \\ & & & \mathfrak{F}_m \end{vmatrix}, \tag{37}$$

([34]) The notation here is different from that in §4.1.

then the l-rank of $\mathfrak{G}$ is equal to the sum $k_1+k_2+\cdots+k_m$ of the l-ranks k_i of $\mathfrak{F}_i$.

This is a generalization of the Frobenius-Schur theorem([35]).

Proof. We can find an l-basis of $\mathfrak{M}(\mathfrak{G})$ such that for a fixed ν the last k_ν basis elements have k_ν l-independent matrices in the place of $\mathfrak{F}_\nu$. After subtracting a suitable linear combination of these basis elements from the first $k-k_\nu$ basis elements, we may assume that the latter have 0 in the place of $\mathfrak{F}_\nu$. On forming the regular representation $\mathfrak{R}$ of $\mathfrak{G}$ with regard to this basis, we obtain

$$\mathfrak{R}=\begin{pmatrix} * & \\ * & \mathfrak{R}_\nu \end{pmatrix}$$

where $\mathfrak{R}_\nu$ is the regular representation of $\mathfrak{F}_\nu$. If f_ν is the degree of $\mathfrak{F}_\nu$, we have $\mathfrak{R}_\nu \simeq (k_\nu/f_\nu)\times\mathfrak{F}_\nu$ (cf. (9.1A)). The constituents $\mathfrak{F}_\nu$ of $\mathfrak{R}$ occupy therefore at least k_ν ordinary rows and columns of $\mathfrak{R}$. The degree k of $\mathfrak{R}$, then, cannot be smaller than the sum of all the k_ν. On the other hand, (37) shows that $k\leqq k_1+\cdots+k_m$, and this proves the statement. At the same time, we see

(10.1B) *The regular representation $\mathfrak{R}$ of $\mathfrak{G}$ in* (10.1A) *contains the constituent $\mathfrak{F}_\nu$ with the multiplicity k_ν/f_ν, where f_ν is the degree of $\mathfrak{F}_\nu$.*

The result (10.1A) can be formulated in the following manner:

(10.1C) *Under the assumption of* (10.1A), *the module $\mathfrak{M}(\mathfrak{G})$ is a direct sum $\mathfrak{M}(\mathfrak{G})=\mathfrak{U}_1\oplus\mathfrak{U}_2\oplus\cdots\oplus\mathfrak{U}_m$ where $\mathfrak{U}_\mu$ consists of those matrices $\mathfrak{M}(\mathfrak{G})$ which have nonzero elements only in the place of the constituent $\mathfrak{M}(\mathfrak{F}_\mu)$ of $\mathfrak{M}(\mathfrak{G})$.*

Proof. Let M_μ be an arbitrary element of $\mathfrak{M}(\mathfrak{F}_\mu)$ $(\mu=1, 2, \cdots, m)$, and set

$$\overline{M}=\begin{bmatrix} M_1 & & & \\ & M_2 & & \\ & & \ddots & \\ & & & M_m \end{bmatrix}.$$

All these $\overline{M}$ form a $(K, \mathfrak{G})$-double module $\overline{\mathfrak{M}}$. We have $\overline{\mathfrak{M}}\supseteq\mathfrak{M}(\mathfrak{G})$, and both these modules have the same l-rank according to (10.1A). Hence $\overline{\mathfrak{M}}=\mathfrak{M}(\mathfrak{G})$. We now choose M_μ arbitrarily in $\mathfrak{M}(\mathfrak{F}_\mu)$, and $M_\nu=0$ for $\nu\neq\mu$. The corresponding $\overline{M}$ form a submodule $\mathfrak{U}_\mu$ of $\mathfrak{M}(\mathfrak{G})$, and $\mathfrak{M}(\mathfrak{G})$ is the direct sum $\mathfrak{U}_1\oplus\cdots\oplus\mathfrak{U}_m$.

2. In order to study further the decomposition into direct sums, we consider two sets of square matrices $\mathfrak{A}$ and $\mathfrak{B}$ of the same degree n, such that $\mathfrak{A}\mathfrak{B}=0$. Let $\mathfrak{V}$ be the space in which the transformations of $\mathfrak{A}$ and $\mathfrak{B}$ take place. Let $\mathfrak{V}_0$ be the subspace consisting of those vectors V_0 for which $\mathfrak{A}V_0=0$. Then we have $\mathfrak{B}\mathfrak{V}\subseteq\mathfrak{V}_0$. If $\mathfrak{V}_0$ has s dimensions and we choose a basis of $\mathfrak{V}$

([35]) Frobenius-Schur [10].

in which the last s vectors form a basis of $\mathfrak{B}_0$, then in the corresponding similar set $P^{-1}\mathfrak{A}P$ the last s columns consist of zeros, and in $P^{-1}\mathfrak{B}P$ the first s rows consist of zeros. Hence

(10.2A) *If $\mathfrak{A}$ and $\mathfrak{B}$ are two sets of square matrices of degree n and $\mathfrak{A}\mathfrak{B}=0$, then we can find a similarity transformation P such that*

$$P^{-1}\mathfrak{A}P = \begin{pmatrix} \mathfrak{A}_1 & \\ \mathfrak{C} & 0 \end{pmatrix}, \qquad P^{-1}\mathfrak{B}P = \begin{pmatrix} 0 & \\ \mathfrak{D} & \mathfrak{B}_1 \end{pmatrix}, \tag{38}$$

where both sets are broken up according to the same scheme $(n-s,\ s \mid n-s,\ s)$.

We may have here $s=0$, if $\mathfrak{B}=0$. Then the second row and column in (38) are missing. Similarly, we may have $s=n$, if $\mathfrak{A}=0$, and then the first row and column in (38) are missing. If $\mathfrak{A}\neq 0$, $\mathfrak{B}\neq 0$, the set $\mathfrak{Q}$ consisting of all sums $A+B$ with A in $\mathfrak{A}$, B in $\mathfrak{B}$ is reducible. This gives

(10.2B) *The set $\mathfrak{U}_\mu$ in (10.1C) cannot be written as a direct sum $\mathfrak{A}\oplus\mathfrak{B}$ with $\mathfrak{A}\neq 0$, $\mathfrak{B}\neq 0$.*

3. As an application of (10.2A), we prove

(10.3A) *Let $\mathfrak{Q}$ be a set of square matrices of degree n which has no constituents 0. If $\mathfrak{Q}$ can be written as a sum $\mathfrak{Q}=\mathfrak{A}\oplus\mathfrak{B}$ with $\mathfrak{A}\neq 0$, $\mathfrak{B}\neq 0$, then there exists a similarity transformation P such that*

$$P^{-1}\mathfrak{Q}P = \begin{pmatrix} \mathfrak{A}_1 & \\ & \mathfrak{B}_1 \end{pmatrix}$$

and $P^{-1}\mathfrak{A}P$ consists of the matrices of $P^{-1}\mathfrak{Q}P$ which have 0 in the place of $\mathfrak{B}_1$ and $P^{-1}\mathfrak{B}P$ consists of those matrices of $P^{-1}\mathfrak{Q}P$ which have 0 in the place of $\mathfrak{A}_1$.

Proof. We may determine P such that $P^{-1}\mathfrak{A}P$ and $P^{-1}\mathfrak{B}P$ have the form (38). The set $\mathfrak{B}_1$ has no constituent 0, since otherwise 0 would also be a constituent of the sum of the two sets (38), and hence of $\mathfrak{Q}$. If $\mathfrak{H}$ is the semigroup generated by $P^{-1}\mathfrak{B}P$, and $\mathfrak{M}(\mathfrak{H})$ its enveloping module, then the matrices M of $\mathfrak{M}(\mathfrak{H})$ are l-annihilators of $P^{-1}\mathfrak{A}P$. Further $\mathfrak{M}(\mathfrak{H})$ breaks up in the same form as $P^{-1}\mathfrak{B}P$ in (38) the first constituent being 0 and the second $\mathfrak{M}(\mathfrak{H}_1)$ where $\mathfrak{H}_1$ is the semigroup generated by $\mathfrak{B}_1$. According to (8.2G) this set $\mathfrak{M}(\mathfrak{H}_1)$ contains the unit matrix I. Let J be a matrix of $\mathfrak{M}(\mathfrak{H})$ which has I in the place of $\mathfrak{M}(\mathfrak{H}_1)$, and let A be an arbitrary element of $\mathfrak{A}$. We set

$$J = \begin{pmatrix} 0 & \\ D & I \end{pmatrix}, \qquad P^{-1}AP = \begin{pmatrix} A_1 & \\ C & 0 \end{pmatrix}.$$

Because of $J(P^{-1}AP)=0$, we have $DA_1+C=0$. We subtract the first row in (38), l-multiplied by D, from the second row and add the second column,

r-multiplied by D, to the first column. This amounts to a similarity transformation (cf. (3.3A)). Afterwards we have $\mathfrak{C}=0$, and we may assume that this is also true in (38). Then $\mathfrak{B}\mathfrak{A}=0$ implies $\mathfrak{D}\mathfrak{A}_1=0$. Since $\mathfrak{A}_1$ also has no constituent 0, it follows that $\mathfrak{D}=0$, and this proves the statement.

Repeated application of (10.3A) gives

(10.3B) *If a set $\mathfrak{Q}$ without constituent* 0 *is a direct sum* $\mathfrak{U}_1\oplus\cdots\oplus\mathfrak{U}_m$ $(\mathfrak{U}_\mu\neq 0)$, *then, after a suitable similarity transformation* P, $P^{-1}\mathfrak{Q}P$ *splits completely into* m *constituents and the matrices of* $P^{-1}\mathfrak{U}_\mu P$ *have coefficients not equal to* 0 *only at the place of the* μ*th of these constituents.*

4. The *radical* $\mathfrak{N}$ of a set $\mathfrak{A}$ of square matrices consists of those matrices N of $\mathfrak{A}$ which are represented by 0 in every irreducible constituent of $\mathfrak{A}$. Then Q is also represented by 0 in the Loewy constituents $L_i(\mathfrak{A})$; i.e., N has zeros in the main diagonal in (14). A simple computation shows that the product of any $L(\mathfrak{A})$ matrices vanishes. If $\mathfrak{A}$ is a ring of matrices, $\mathfrak{N}$ is a nilpotent ideal, $\mathfrak{N}^L=0$, for $L=L(\mathfrak{A})$.

We can easily study the radical of the enveloping module $\mathfrak{M}(\mathfrak{G})$ of a semi-group $\mathfrak{G}$, provided that $\mathfrak{G}$ has been brought into a suitable form by a similarity transformation.

(10.4A) *Let $\mathfrak{G}$ be a semi-group which splits into irreducible constituents*

$$\mathfrak{G}=\begin{pmatrix}\mathfrak{F}_1 & & \\ & \ddots & \\ * & & \mathfrak{F}_m\end{pmatrix},\qquad \mathfrak{M}(\mathfrak{G})=\begin{pmatrix}\mathfrak{M}(\mathfrak{F}_1) & & \\ & \ddots & \\ & & \mathfrak{M}(\mathfrak{F}_m)\end{pmatrix}. \tag{39}$$

Then the radical $\mathfrak{N}$ of $\mathfrak{M}(\mathfrak{G})$ has at least the l-rank $k-\lambda$ where k is the l-rank of $\mathfrak{G}$ and λ the degree of the first Loewy constituent $\mathfrak{L}_1(\mathfrak{R})$ of the regular representation $\mathfrak{R}$ of $\mathfrak{G}$.

Proof. Let $M_1,\cdots,M_k$ be an l-basis of $\mathfrak{M}(\mathfrak{G})$ with regard to which the regular representation $\mathfrak{R}$ appears in its lower Loewy normal form. If G is an arbitrary element of $\mathfrak{G}$, we have

$$M_\kappa G=\sum r_{\kappa\lambda}M_\lambda$$

where $R=(r_{\kappa\lambda})$ is the matrix of $\mathfrak{R}$, associated with G. If $\mathfrak{B}$ is one of the $\mathfrak{F}_\kappa$, and M_κ corresponds to U_κ in $\mathfrak{M}(\mathfrak{B})$ and G corresponds to B, we have

$$U_\kappa B=\sum r_{\kappa\lambda}U_\lambda.$$

We now apply (7.8A) setting $U_\kappa=(h_{\alpha\beta}^{(\kappa)})$. Then $P_\mu=(h_{\mu\beta}^{(\alpha)})$ will intertwine $\mathfrak{R}$ and $\mathfrak{B}$. Because of (5.2A), only the last λ rows of P_μ contain coefficients not equal to 0. Hence

$$h_{\mu\beta}^{(\alpha)}=0 \qquad \text{for } \alpha\leqq k-\lambda.$$

This shows $U_\kappa=0$ for $\kappa \leqq k-\lambda$. Hence $M_1, \cdots, M_{k-\lambda}$ are represented by 0 in each $\mathfrak{M}(\mathfrak{F}_\kappa)$ and, therefore, belong to $\mathfrak{N}$.

(10.4B) *If the semigroup $\mathfrak{G}$ splits into irreducible constituents and the radical of $\mathfrak{M}(\mathfrak{G})$ vanishes, then $\mathfrak{G}$ is completely reducible.*

Proof. We have here $k=\lambda$; i.e., $\mathfrak{N}$ is completely reducible, $L(\mathfrak{N})=1$. We denote by $\overline{\mathfrak{G}}$ the set obtained from $\mathfrak{G}$ by replacing everything below the main diagonal in (39) by 0's, and omitting all constituents 0. According to (8.2G), a suitable linear combination of the elements of $\overline{\mathfrak{G}}$ is equal to the unit matrix. A corresponding linear combination of the elements of $\mathfrak{G}$ gives a matrix J of $\mathfrak{M}(\mathfrak{G})$ which in (39) has a unit matrix in the place of every $\mathfrak{M}(\mathfrak{F}_\rho)\neq 0$ and, of course, 0 in the place of every $\mathfrak{M}(\mathfrak{F}_\rho)=0$. The product JG of J with an element G of $\mathfrak{G}$ has the same main diagonal as G. Then $G-JG$ lies in the radical of $\mathfrak{M}(\mathfrak{G})$, and hence $JG=G$. Now (8.2F) can be applied. We obtain $L(\mathfrak{G})=L(\mathfrak{N})=1$, i.e., $\mathfrak{G}$ is completely reducible.

If K is noncommutative, the converse of assertion (10.4B) need not be true.

5. Repeated application of (10.2A) now gives

(10.5A) *If a set $\mathfrak{Q}$ of square matrices is a sum of sets $\mathfrak{Q}_1, \mathfrak{Q}_2, \cdots, \mathfrak{Q}_r$, if $\mathfrak{Q}_i\mathfrak{Q}_j=0$ for $i<j$, and if no $\mathfrak{Q}_i$ lies in the radical of $\mathfrak{Q}$, then, after a similarity transformation, $\mathfrak{Q}$ will split into r constituents $\mathfrak{T}_1, \mathfrak{T}_2, \cdots, \mathfrak{T}_r$. The matrices of $\mathfrak{Q}_\rho$ have 0 in the place of every $\mathfrak{T}_\sigma$, $\sigma\neq\rho$.*

Proof. We apply (10.2A) to the case that $\mathfrak{A}$ is the sum of $\mathfrak{Q}_1, \cdots, \mathfrak{Q}_{r-1}$ and $\mathfrak{B}=\mathfrak{Q}_r$. We then have an equation (38). Here, $\mathfrak{B}_1\neq 0$, since $\mathfrak{Q}_r$ does not lie in the radical of $\mathfrak{Q}$. Let $\mathfrak{Q}_i^*$ be the set which stands in $\mathfrak{Q}_i\subseteq\mathfrak{Q}$ in the place of $\mathfrak{A}_1$ ($i=1, 2, \cdots, r-1$). Then $\mathfrak{A}_1$ is the sum of $\mathfrak{Q}_1^*, \cdots, \mathfrak{Q}_{r-1}^*$, and $\mathfrak{Q}_i^*$ does not belong to the radical of $\mathfrak{A}_1$, since otherwise $\mathfrak{Q}_i$ would belong to the radical of $\mathfrak{Q}$. If the theorem is true for the sums of $r-1$ sets, it now follows for the sum of r sets.

11. Rings which contain $n\times K$

1. We now consider rings of matrices $\mathfrak{A}$ of degree n with coefficients in the division ring K which are at the same time K-left modules and K-right modules, i.e., which contain γA and $A\gamma$ for all A in $\mathfrak{A}$ and all γ in K. Of course, this property will not always be preserved under similarity transformations of $\mathfrak{A}$.

If $\mathfrak{A}$ is a ring which is a K-left module, we have $\mathfrak{M}(\mathfrak{A})=\mathfrak{A}$ in the notation of §8.1. If $\mathfrak{A}$ has no constituent 0, then $\mathfrak{A}$ contains the unit matrix according to (8.2G), and hence all the matrices γI, γ in K. These matrices form a set $\mathfrak{K}$ isomorphic with K which we may denote by $n\times K$, if we identify the matrix (γ) of first degree with γ. Any ring $\mathfrak{A}$ which contains $\mathfrak{K}=n\times K$ is a K-left module and a K-right module.

We prove several lemmas which connect $\mathfrak{A}$ with sets of matrices whose coefficients lie in the centre Z of K. This centre Z is a field.

(11.1A) *If $\mathfrak{A}$ is a ring of matrices which is a K-left module and a K-right module, then an l-basis $A_1, A_2, \cdots, A_k$ can be chosen such that the coefficients of each A_κ lie in the centre Z of K. The A_κ form a basis of the algebra $\overline{\mathfrak{A}} = \mathfrak{A} \cap [Z]_n$ over the field Z. We have $\mathfrak{A} = \mathfrak{M}(\overline{\mathfrak{A}})$.*

Proof. The set $\mathfrak{A}$ obviously is a $(K, \mathfrak{K})$-double module. Now $\mathfrak{K} = n \times K$ is completely reducible with K as its only irreducible constituent. According to (7.6A), the same is true for the set $\mathfrak{K}^*$ which $\mathfrak{A}$ associates with $\mathfrak{K}$. If we choose a suitable l-basis A_κ in $\mathfrak{A}$, we have $\mathfrak{K}^* = k \times K$ where k is the l-rank of $\mathfrak{A}$. Then $A_\kappa \gamma = \gamma A_\kappa$ for every γ in K. This shows that the coefficients of A_κ lie in Z.

Every element A of $\mathfrak{A}$ has the form $A = \sum \gamma_\kappa A_\kappa$ with coefficients γ_κ in K. These γ_κ are uniquely determined, and we have a system of n^2 linear equations for them. If A belongs to $[Z]_n$, i.e., if the coefficients of A lie in Z, then the coefficients of these linear equations lie in Z. Hence (cf. §3.4) the γ_κ themselves lie in Z. This proves (11.1A).

We now consider the commuting ring $\mathfrak{C}(\mathfrak{A})$. We prove

(11.1B) *If $\mathfrak{A}$ is a ring of matrices which is a K-left module and has no constituent 0, then $\mathfrak{C}(\mathfrak{A}) = \mathfrak{C}(\overline{\mathfrak{A}}) \cap [Z]_n$ and $\mathfrak{C}(\overline{\mathfrak{A}}) = \mathfrak{M}(\mathfrak{C}(\mathfrak{A}))$.*

Proof. Here, $\mathfrak{K} = n \times K \subseteq \mathfrak{A}$ and hence $\mathfrak{C}(\mathfrak{A}) \subseteq \mathfrak{C}(n \times K) = [\mathfrak{C}(K)]_n = [Z]_n$. Further $\mathfrak{C}(\mathfrak{A}) \subseteq \mathfrak{C}(\overline{\mathfrak{A}})$. On the other hand, every matrix M of the intersection $\mathfrak{C}(\overline{\mathfrak{A}}) \cap [\mathfrak{Z}]_n$ commutes with the A_κ of (11.1A) and with all γ in K. Hence M belongs to $\mathfrak{C}(\mathfrak{A})$, and $\mathfrak{C}(\mathfrak{A}) = \mathfrak{C}(\overline{\mathfrak{A}}) \cap [Z]_n$. The ring $\mathfrak{C}(\overline{\mathfrak{A}})$ contains $\mathfrak{K}$. If we apply (11.1A) to it, we obtain $\mathfrak{C}(\overline{\mathfrak{A}}) = \mathfrak{M}(\mathfrak{C}(\mathfrak{A}))$.

2. (11.2A) *If $\mathfrak{A}$ is a set of matrices of degree n which contains $\mathfrak{K} = n \times K$, we may determine a matrix P with coefficients in the centre Z of K, such that $P^{-1}\mathfrak{A}P = \mathfrak{A}^*$ splits into irreducible constituents. If $\mathfrak{A}$ is completely reducible, we may add here the additional condition that $\mathfrak{A}$ splits completely into irreducible constituents.*

Proof. We split $\mathfrak{A}$ into irreducible constituents using a similarity transformation Q with coefficients in K,

$$Q^{-1}\mathfrak{A}Q = \begin{pmatrix} \mathfrak{A}_1 & & \\ & \ddots & \\ * & & \mathfrak{A}_m \end{pmatrix}. \tag{40}$$

If $\mathfrak{A}$ is completely reducible, we may assume that all the terms below the main diagonal vanish. The subset $Q^{-1}\mathfrak{K}Q$ of $Q^{-1}\mathfrak{A}Q$ is completely reducible, and K is its only irreducible constituent. If we use (40) only for $Q^{-1}\mathfrak{K}Q$, the set $\mathfrak{K}_\mu$

which takes the place of $\mathfrak{A}_\mu$ is completely reducible, and K is its only irreducible constituent. After applying a suitable similarity transformation to (40), we may assume that $\mathfrak{K}_\mu = f_\mu \times K$ where f_μ is the degree of $\mathfrak{A}_\mu$. Now the set $Q^{-1}\mathfrak{K}Q$ splits into n constituents K. Since it is completely reducible, we can transform it into $n \times K$ by elementary similarity transformations(26), §4.5. If we apply these elementary similarity transformations to (40), the triangular form will not be changed. We may therefore assume right from the beginning that $Q^{-1}\mathfrak{K}Q = n \times K = \mathfrak{K}$. If $\mathfrak{A}$ is completely reducible, no elementary similarity transformations are needed. We now have $Q^{-1}(\gamma I)Q = \gamma I$ for every γ in K. Then $\gamma Q = Q\gamma$, i.e., Q has coefficients in Z, and we may take $P = Q$.

3. Let $\mathfrak{A}$ be a ring of matrices of degree n which contains $n \times K$. If $\mathfrak{B}$ is a homomorphic set of matrices of degree m, and if the element γI_n of $\mathfrak{A}$ corresponds to γI_m in $\mathfrak{B}$ for every γ in K, then $\mathfrak{B}$ is said to be a *representation* of degree m of $\mathfrak{A}$. If we split $\mathfrak{A}$ into irreducible constituents by means of the transformation P of (11.2A), the irreducible constituents of $\mathfrak{A}$ will then be representations of $\mathfrak{A}$.

If we use the basis A_κ of (11.1A) for the definition of the regular representation $\mathfrak{R}$ of $\mathfrak{A}$, then $\mathfrak{R}$ will actually be a representation of $\mathfrak{A}$.

Any two representations $\mathfrak{B}_1$ and $\mathfrak{B}_2$ of $\mathfrak{A}$ are to be considered as related sets (§3.1) with $\mathfrak{Z} = \mathfrak{A}$. If $\mathfrak{B}_1$ and $\mathfrak{B}_2$ are similar, say $\mathfrak{B}_1 = Q^{-1}\mathfrak{B}_2 Q$, then $\gamma I = Q^{-1}(\gamma I)Q$ for every γ in K. This implies that Q has coefficients in Z.

(11.3A) *If two representations of the ring $\mathfrak{A} \supseteq n \times K$ are similar, then the corresponding similarity transformation has coefficients in the centre of K.*

4. We now derive the results of the structure theory of algebras(27).

(11.4A) *If $\mathfrak{A} \neq 0$ is an irreducible ring of matrices which is a K-left module, then $\mathfrak{A} \sim [\mathfrak{T}]_s$ where $\mathfrak{T}$ is a division ring consisting of matrices and $s > 0$ an integer. We have $\mathfrak{C}(\mathfrak{C}(\mathfrak{A})) = \mathfrak{A}$.*

Proof. Since $\mathfrak{A}$ has no constituent 0, we have $\mathfrak{K} = n \times K \subseteq \mathfrak{A}$. Obviously, $\mathfrak{C}(\mathfrak{C}(\mathfrak{A})) \supseteq \mathfrak{A} \supseteq \mathfrak{K}$. On applying (11.1A) to this ring $\mathfrak{C}(\mathfrak{C}(\mathfrak{A}))$ we see that it has a basis consisting of matrices C_ρ with coefficients in Z. These matrices C_ρ have the following two properties: (a) they belong to $[Z]_n$; (b) they commute with every element of $\mathfrak{C}(\bar{\mathfrak{A}}) \cap [Z]_n$, which is equal to $\mathfrak{C}(\mathfrak{A})$ because of (11.1B).

From (11.1A) it follows that $\bar{\mathfrak{A}}$ is irreducible with regard to Z. Let us consider for the moment only matrices with coefficients in Z. Then (9.4A) shows that the commuting ring of the commuting ring of $\bar{\mathfrak{A}}$ is $\bar{\mathfrak{A}}$ itself. In other words: every matrix C with the properties (a) and (b) belongs to $\bar{\mathfrak{A}}$. Then the C_ρ belong to $\bar{\mathfrak{A}} \subseteq \mathfrak{A}$ and hence $\mathfrak{C}(\mathfrak{C}(\mathfrak{A})) \subseteq \mathfrak{A}$ which implies $\mathfrak{C}(\mathfrak{C}(\mathfrak{A})) = \mathfrak{A}$.

We can now use the argument of §9.3. We set $\mathfrak{B} = \mathfrak{C}(\mathfrak{A})$; this set is com-

(26) The degrees n_λ in §3.3 are here to be taken as equal to 1.

(27) Cf., for instance, Albert [1, 2], Deuring [7].

pletely reducible and has only one irreducible constituent $\mathfrak{W}$. We may set

$$Q^{-1}\mathfrak{B}Q = s \times \mathfrak{W}$$

where Q is a matrix with coefficients in K (not necessarily in Z), and $\mathfrak{W}$ is irreducible and a division ring. Then

$$Q^{-1}\mathfrak{A}Q = \mathfrak{C}(s \times \mathfrak{W}) = [\mathfrak{C}(\mathfrak{W})]_s,$$

and $\mathfrak{C}(\mathfrak{W}) = \mathfrak{T}$ itself is irreducible, and a division ring. This proves (11.4A).

We now prove easily in the familiar manner that the ring $\mathfrak{A}$ is simple, i.e., possesses no proper subideal. There is no properly nilpotent element not equal to 0 in $\mathfrak{A}$.

Consider an arbitrary ring $\mathfrak{A}$ of matrices which contains $n \times K$. We determine a similarity transformation with the properties stated in (11.2A). Since the elements of $n \times K$ are transformed into themselves, we may assume without restriction that $\mathfrak{A}$ itself splits into irreducible constituents,

$$\mathfrak{A} = \begin{pmatrix} \mathfrak{A}_1 & & \\ & \ddots & \\ * & & \mathfrak{A}_r \end{pmatrix}. \tag{41}$$

Using (11.3A), we easily see that we may assume that similar $\mathfrak{A}_\rho$ are always equal. Let $\mathfrak{F}_1, \mathfrak{F}_2, \cdots, \mathfrak{F}_m$ be the distinct irreducible constituents appearing, and denote the l-rank of $\mathfrak{F}_\mu$ by k_μ. Then $\mathfrak{M}(\mathfrak{A}) = \mathfrak{A}$, $\mathfrak{M}(\mathfrak{F}_\mu) = \mathfrak{F}_\mu$.

If we replace everything below the main diagonal in (41) by 0, we obtain a representation $\mathfrak{A}^*$ of $\mathfrak{A}$. The elements of the radical $\mathfrak{N}$ of $\mathfrak{A}$ and only these are represented by 0; we see that $\mathfrak{A}/\mathfrak{N}$ and $\mathfrak{A}^*$ are isomorphic. From (10.1A) it follows that $\mathfrak{A}^*$ has the l-rank $\sum k_\mu$. Hence

(11.4B) *If $\mathfrak{A}$ is a ring of matrices which contains $n \times K$, its l-rank is given by $k = k_1 + k_2 + \cdots + k_m + \nu$ where $k_1, k_2, \cdots, k_m$ are the l-ranks of the non-similar irreducible constituents $\mathfrak{F}_1, \mathfrak{F}_2, \cdots, \mathfrak{F}_m$ of $\mathfrak{A}$ and ν is the l-rank of the radical $\mathfrak{N}$ of $\mathfrak{A}$. If in each $\mathfrak{F}_\mu$ an arbitrary element F_μ has been chosen, then there are elements A of $\mathfrak{A}$ which are represented by F_μ in $\mathfrak{F}_\mu$ for $\mu = 1, 2, \cdots, m$.*

If $\mathfrak{A}$ is completely reducible, we may assume that (41) splits completely into irreducible constituents. We then find $\mathfrak{N} = 0$. Conversely, if $\mathfrak{N} = 0$, it follows from (10.4B) that $\mathfrak{A}$ is completely reducible. A ring is semisimple, if its radical vanishes. Hence

(11.4C) *A ring $\mathfrak{A} \supseteq n \times K$ is semisimple, if and only if $\mathfrak{A}$ is completely reducible.*

Ordinarily, the radical is defined as the set of all properly nilpotent elements N of $\mathfrak{A}$. But to such an N, there corresponds a properly nilpotent N_ρ of $\mathfrak{F}_\rho$. Since $\mathfrak{F}_\rho$ is irreducible, we have $N_\rho = 0$. Hence N belongs to $\mathfrak{N}$. Con-

versely, every element of $\mathfrak{N}$ is properly nilpotent. Both definitions of the radical coincide.

The rings $\mathfrak{A}/\mathfrak{N}$ and $\mathfrak{A}^*$ were isomorphic. Hence

(11.4D) *If the ring* $\mathfrak{A}\supseteq n\times K$ *has the radical* $\mathfrak{N}$, *then* $\mathfrak{A}/\mathfrak{N}$ *is semisimple.*

If $\mathfrak{A}$ is semisimple, we can apply (10.1C) and find:

(11.4E) *Let* $\mathfrak{A}\supseteq n\times K$ *be a semisimple ring. If* $\mathfrak{F}_1, \mathfrak{F}_2, \cdots, \mathfrak{F}_m$ *are the non-similar irreducible constituents of* $\mathfrak{A}$, *then* $\mathfrak{A}$ *is the direct sum* $\mathfrak{A}=\mathfrak{U}_1\oplus\mathfrak{U}_2\oplus\cdots\oplus\mathfrak{U}_m$ *of m simple rings and* $\mathfrak{U}_\mu$ *is isomorphic with* $\mathfrak{F}_\mu$.

On combining the last part of (11.4A) with §6.2, we obtain

(11.4F) *If* $\mathfrak{A}$ *is a completely reducible ring which contains* $n\times K$, *then* $\mathfrak{C}(\mathfrak{C}(\mathfrak{A}))=\mathfrak{A}$.

Finally, we can show that

(11.4G) *If* $\mathfrak{B}$ *is a simple ring of matrices, and* $\mathfrak{B}\supseteq n\times K$, *then* $\mathfrak{B}\sim t\times\mathfrak{A}$, *where* $t>0$ *is an integer and* $\mathfrak{A}$ *an irreducible ring. Then* $\mathfrak{B}$ *is isomorphic to the ring* $\mathfrak{A}$ *whose structure is described by* (11.4A).

Proof. The radical of $\mathfrak{B}$ must vanish. Therefore, $\mathfrak{B}$ is completely reducible. From (11.4E) it follows that $\mathfrak{B}$ has only one irreducible constituent.

5. We consider an arbitrary ring $\mathfrak{A}$ of matrices which contains $n\times K$ and a representation $\mathfrak{B}$ of $\mathfrak{A}$. Let $A_1, \cdots, A_k$ be an l-basis of $\mathfrak{A}$ and A an arbitrary element of $\mathfrak{A}$. The regular representation $\mathfrak{R}$ is defined by

$$A_\kappa A = \sum_\lambda r_{\kappa\lambda}A_\lambda, \qquad r_{\kappa\lambda} \text{ in } K. \tag{42}$$

If $A_\kappa\rightarrow B_\kappa$, $A\rightarrow B$ are the associated elements in $\mathfrak{B}$, we find

$$B_\kappa B_\kappa = \sum r_{\kappa\lambda}B_\lambda. \tag{43}$$

We may assume that for a certain t the elements $B_1=\cdots=B_t=0$ and that $B_{t+1}, \cdots, B_k$ are l-independent. On comparing (42) and (43), we see that the regular representation of $\mathfrak{B}$ appears as an end constituent of $\mathfrak{R}$. Using (8.2A) we now find:

(11.5A) *Let* $\mathfrak{A}$ *be a ring of matrices containing* $n\times K$. *Every representation* $\mathfrak{B}$ *of* $\mathfrak{A}$ *of degree m appears as an end constituent of* $m\times\mathfrak{R}$ *where* $\mathfrak{R}$ *is the regular representation of* $\mathfrak{A}$. *Further,* $\mathfrak{B}$ *appears as a constituent of* $mn\times\mathfrak{A}$.

As corollaries, we obtain:

(11.5B) *If* $\mathfrak{B}$ *is a representation of* $\mathfrak{A}$, *then* $L(\mathfrak{B})\subseteq L(\mathfrak{A})$.

(11.5C) *Every irreducible representation of* $\mathfrak{A}$ *appears as a constituent of* $\mathfrak{A}$.

The following theorems are sometimes useful.

(11.5D) *If $\mathfrak{B}_1$ and $\mathfrak{B}_2$ are two representations of $\mathfrak{A}$ which have no irreducible constituent in common, then we can find an element Q of $\mathfrak{A}$ which is represented by the unit matrix in $\mathfrak{B}_1$ and by 0 in $\mathfrak{B}_2$.*

Proof. According to (11.4B), we can find an element A of $\mathfrak{A}$ which is represented by the unit matrix in every irreducible constituent of $\mathfrak{B}_1$ and by 0 in every irreducible constituent of $\mathfrak{B}_2$. Then A corresponds to a radical element B_2 of $\mathfrak{B}_2$. If we replace A by a power of A, we may assume $B_2=0$. If A is represented by B_1 in $\mathfrak{B}_1$, then B_1-I lies in the radical of $\mathfrak{B}_1$; we have $(B_1-I)^t=0$ for some integer $t>0$. Hence we may write I as a polynomial $f(B_1)$ without constant term of B_1. Then $Q=f(A)$ will satisfy the required conditions.

(11.5E) *If $\mathfrak{B}_1$ and $\mathfrak{B}_2$ are two representations of $\mathfrak{A}$ which have no irreducible constituent in common, and if B_1 and B_2 are arbitrary elements of $\mathfrak{B}_1$ and $\mathfrak{B}_2$ respectively, then we may find an element A of $\mathfrak{A}$ which is represented by B_1 in $\mathfrak{B}_1$ and by B_2 in $\mathfrak{B}_2$.*

Proof. Let $A^{(1)}$ be an element of $\mathfrak{A}$ which is represented by B_1 in $\mathfrak{B}_1$ and determine Q as in (11.5D). Then $QA^{(1)}$ is represented by B_1 in $\mathfrak{B}_1$ and by 0 in $\mathfrak{B}_2$. Similarly, we may find an element $\overline{Q}A^{(2)}$ of $\mathfrak{A}$ which is represented by 0 in $\mathfrak{B}_1$ and by B_2 in $\mathfrak{B}_2$. Then we may set $A=QA^{(1)}+\overline{Q}A^{(2)}$.

(11.5F) *If $\mathfrak{B}$ is a representation of $\mathfrak{A}$, the radical of $\mathfrak{A}$ is represented by the radical of $\mathfrak{B}$.*

Proof. It is clear that radical elements of $\mathfrak{A}$ are represented by radical elements of $\mathfrak{B}$. Conversely, let B be a radical element of $\mathfrak{B}$. We set $\mathfrak{B}_1=\mathfrak{B}$; for $\mathfrak{B}_2$ we take the representation of $\mathfrak{A}$ which splits completely into those irreducible representations of $\mathfrak{A}$ which do not appear in $\mathfrak{B}$. We then apply (11.5E) to the case $B_1=B$, $B_2=0$. The corresponding A lies in the radical of $\mathfrak{A}$ and is represented by B in $\mathfrak{B}$.

12. The regular representation of rings which contain $n\times K$

1. We consider again a ring $\mathfrak{A}$ of matrices which is a K-left module, in particular a ring $\mathfrak{A}$ which contains $n\times K$. The regular representation $\mathfrak{R}$ of $\mathfrak{A}$ is a set of linear transformations of $\mathfrak{A}$ where $\mathfrak{A}$ is considered as a contragredient vector space. The element A of $\mathfrak{A}$ is associated with the transformation $R(A)$ which maps the variable element X of $\mathfrak{A}$ upon $XA=X^*$. In particular, if $A_1, A_2, \cdots, A_k$ is an l-basis of $\mathfrak{A}$, we have

$$A_\kappa^* = A_\kappa A = \sum r_{\kappa\lambda}A_\lambda$$

where $R(A)=(r_{\kappa\lambda})$.

A subspace $\mathfrak{T}$ of $\mathfrak{A}$ which is invariant under the transformation of $\mathfrak{R}$, then, is a right ideal $\mathfrak{T}$ of $\mathfrak{A}$ which is a K-left module. Since we shall consider the elements γ of K as operators of $\mathfrak{A}$, where the operation is defined as l-multi-

plication by γ, we shall tacitly assume that the right ideals considered are K-left modules.

Any splitting of $\mathfrak{R}$ into constituents will correspond to an ascending chain of invariant subspaces $\mathfrak{T}$, (cf. §4.4), i.e., an ascending chain of r-ideals of $\mathfrak{A}$. More explicitly, if

$$(44) \qquad \mathfrak{R} = \begin{pmatrix} \mathfrak{H}_1 & & \\ \cdot & \ddots & \\ & & \mathfrak{H}_m \end{pmatrix}$$

where $\mathfrak{H}_\mu$ is a constituent of degree h_μ, then the linear combinations $\sum \gamma_\nu A_\nu$ of the first $h_1+h_2+\cdots+h_\rho$ basis elements with coefficients γ_ν in K form an r-ideal $\mathfrak{T}_\rho$ of $\mathfrak{A}$. We set $\mathfrak{T}_0=(0)$ and have

$$(45) \qquad \mathfrak{T}_0 = (0) \subset \mathfrak{T}_1 \subset \mathfrak{T}_2 \subset \cdots \subset \mathfrak{T}_{m-1} \subset \mathfrak{T}_m = \mathfrak{A}.$$

Conversely, assume that such a chain of r-ideals is given where $\mathfrak{T}_\mu$ has the l-rank t_μ. We choose the l-basis $A_1, A_2, \cdots, A_k$ of $\mathfrak{A}$ such that the first t_μ basis elements form an l-basis of $\mathfrak{T}_\mu$ for $\mu=1, 2, \cdots, m$. Then the regular representation $\mathfrak{R}$ formed with regard to this basis A_κ breaks up in the form (44), the degrees h_μ being given by $h_\mu=t_\mu-t_{\mu-1}$. We say that the l-basis A_κ has been *adapted to the chain* (45) of r-ideals. If we change the A_κ corresponding to $\mathfrak{H}_\mu$, i.e., the A_κ with $t_{\mu-1}<\kappa\leqq t_\mu$, in such a manner that the new basis is still adapted to the chain (45), then $\mathfrak{R}$ undergoes a similarity transformation of the type (3.3C).

2. We assume that the ring $\mathfrak{A}$ contains $n\times K$. Every r-ideal $\mathfrak{T}$ is a K-left module and a K-right module. Then, (11.1A) can be applied. The set $\overline{\mathfrak{T}}=\mathfrak{T}\cap[Z]_n$ will be a right ideal of $\overline{\mathfrak{A}}=\mathfrak{A}\cap[Z]_n$ considered as an algebra over Z. Every right ideal $\overline{\mathfrak{T}}$ of $\overline{\mathfrak{A}}$ will be obtained in this form, if we take $\mathfrak{T}=\mathfrak{M}(\overline{\mathfrak{T}})$.

(12.2A) *If $\mathfrak{A}$ is a ring containing $n\times K$, then by*

$$\overline{\mathfrak{T}} = \mathfrak{T} \cap [Z]_n, \qquad \mathfrak{T} = \mathfrak{M}(\overline{\mathfrak{T}})$$

there is defined a (1-1) *correspondence between the set of the r-ideals $\mathfrak{T}$ of $\mathfrak{A}$ and the set of the r-ideals $\overline{\mathfrak{T}}$ of $\overline{\mathfrak{A}}=\mathfrak{A}\cap[Z]_n$. Here $\overline{\mathfrak{A}}$ is considered as an algebra over the centre Z of K.*

Further, we easily see from (11.1A) that

(12.2B) *If an ascending chain of r-ideals of $\mathfrak{A}$ is given, we can choose an l-basis A_κ of $\mathfrak{A}$ which is adapted to this basis such that every A_κ has coefficients in the centre Z.*

(12.2C) *Let $\overline{A}_1, \overline{A}_2, \cdots, \overline{A}_k$ be an l-basis of $\overline{\mathfrak{A}}$ such that the regular representation $\overline{\mathfrak{R}}$ of $\overline{\mathfrak{A}}$ formed with regard to this basis splits into constituents which are*

irreducible in Z. If the same basis $\bar{A}_\kappa$ is used for the definition of the regular representation $\mathfrak{R}$ of $\mathfrak{A}$, then $\mathfrak{R}$ splits into constituents which are irreducible in K.

3. We now discuss conditions under which the $\mathfrak{H}_\nu$ in (44) are completely reducible.

(12.3A) *Let $\mathfrak{A}$ be a ring containing $n\times K$ whose radical is $\mathfrak{N}$. Let $(0)=\mathfrak{T}_0\subset\mathfrak{T}_1\subset\cdots\subset\mathfrak{T}_m=\mathfrak{A}$ be a chain of right ideals. In the corresponding splitting* (44) *of the regular representation $\mathfrak{R}$, the constituent $\mathfrak{H}_\mu$ is completely reducible, if and only if $\mathfrak{T}_\mu\mathfrak{N}\subseteq\mathfrak{T}_{\mu-1}$.*

Proof. If $\mathfrak{T}_\mu\mathfrak{N}\subseteq\mathfrak{T}_{\mu-1}$, then $\mathfrak{N}$ will be represented by 0 in $\mathfrak{H}_\mu$. From (11.5F) and (11.4C) it follows that $\mathfrak{H}_\mu$ has the radical (0) and hence is completely reducible. Conversely, if $\mathfrak{H}_\mu$ is completely reducible, it represents $\mathfrak{N}$ by 0. Then $\mathfrak{T}_\mu\mathfrak{N}\subseteq\mathfrak{T}_{\mu-1}$, as was stated.

The complete reducibility of $\mathfrak{H}_\mu$ is, of course, equivalent to the complete reducibility of $\mathfrak{T}_\mu/\mathfrak{T}_{\mu-1}$ considered as an additive group with the elements of K as l-operators and the elements of $\mathfrak{A}$ as r-operators.

In order to obtain the lower Loewy normal form of $\mathfrak{R}$, we must choose $\mathfrak{T}_{m-1}$ as small as possible such that $\mathfrak{T}_m\mathfrak{N}\subseteq\mathfrak{T}_{m-1}$; then $\mathfrak{T}_{m-2}$ as small as possible such that $\mathfrak{T}_{m-1}\mathfrak{N}\subseteq\mathfrak{T}_{m-2}$. Thus

(12.3B) *Let $\mathfrak{A}$ be a ring containing $n\times K$ whose radical is $\mathfrak{N}$. The lower Loewy normal form of the regular representation $\mathfrak{R}$ of $\mathfrak{A}$ corresponds to the chain of r-ideals $(0)=\mathfrak{N}^L\subset\mathfrak{N}^{L-1}\subset\mathfrak{N}^{L-2}\subset\cdots\subset\mathfrak{N}\subset\mathfrak{N}^0=\mathfrak{A}$*(38)*. The exponent L here is equal to the number $L(\mathfrak{A})=L(\mathfrak{R})$ of Loewy constituents of $\mathfrak{A}$ and $\mathfrak{R}$* (cf. (8.2B)).

Similarly, we obtain from (12.3A) the theorem that

(12.3C) *Under the assumptions of* (12.3B), *the upper Loewy normal form of $\mathfrak{R}$ corresponds to the series of ideals $(0)=\mathfrak{Q}_0\subset\mathfrak{Q}_1\subset\cdots\subset\mathfrak{Q}_L=\mathfrak{A}$, where $\mathfrak{Q}_i$ consists of the l-annihilators of $\mathfrak{N}^i$ in $\mathfrak{A}$.*

If we consider $\mathfrak{A}$ as an additive group with the elements of K as l-operators and the elements of $\mathfrak{A}$ as r-operators, we may say that $\mathfrak{N}^L\subset\mathfrak{N}^{L-1}\subset\cdots\subset\mathfrak{A}$ and $\mathfrak{Q}_0\subset\mathfrak{Q}_1\subset\cdots\subset\mathfrak{Q}_L$ are the upper and the lower Loewy series of $\mathfrak{A}$.

4. From (12.3B) we see that the degree λ of $\mathfrak{L}_1(\mathfrak{R})$ is equal to $k-\nu$ where k is the l-rank of $\mathfrak{A}$ and ν is the l-rank of $\mathfrak{N}$. In the notation of (11.4B) $\lambda=\sum k_\mu$. But the argument of §10.1 easily shows that every $\mathfrak{F}_\mu$ appears at least k_μ/f_μ times in $\mathfrak{L}_1(\mathfrak{R})$. Therefore, the constituents $\mathfrak{F}_\mu$ occupy at least k_μ ordinary rows of $\mathfrak{L}_1(\mathfrak{R})$. Because $\lambda=\sum k_\mu$, we obtain:

(12.4A) *If $\mathfrak{F}_\mu$ is an irreducible representation of $\mathfrak{A}$ of l-rank k_μ and of degree f_μ, then $\mathfrak{F}_\mu$ appears exactly k_μ/f_μ times in the first Loewy constituent $\mathfrak{L}_1(\mathfrak{R})$ of the regular representation $\mathfrak{R}$ of $\mathfrak{A}$.*

(38) Cf. Nesbitt [19].

Bibliography

1. A. A. Albert, *Modern Higher Algebra*, Chicago, 1937.

2. ———, *Structure of Algebras*, American Mathematical Society Colloquium Publications, vol. 24, 1939.

3. R. Brauer, *Ueber Systeme hyperkomplexer Zahlen*, Mathematische Zeitschrift, vol. 29 (1929), pp. 79–107.

4. R. Brauer and C. Nesbitt, *On the regular representation of algebras*, Proceedings of the National Academy of Sciences, vol. 23 (1937), pp. 236–240.

5. W. Burnside, *Reducibility of any group of linear substitutions*, Proceedings of the London Mathematical Society, (2), vol. 3 (1905), pp. 430–434.

6. A. H. Clifford, *Representations induced in an invariant subgroup*, Annals of Mathematics, (2), vol. 38 (1937), pp. 533–550.

7. M. Deuring, *Algebren*, Ergebnisse der Mathematik, vol. 4, 1935.

8. H. Fitting, *Die Theorie der Automorphismenringe Abelscher Gruppen und ihr Analogon bei nichtkommutativen Gruppen*, Mathematische Annalen, vol. 107 (1932), pp. 514–542.

9. G. Frobenius, *Theorie der hyperkomplexen Grössen*, Sitzungsberichte der Preussischen Akademie der Wissenschaften, 1903, pp. 504–537.

10. G. Frobenius and I. Schur, *Ueber die Aequivalenz von Gruppen linearer Substitutionen*, Sitzungsberichte der Preussischen Akademie der Wissenschaften, 1906, pp. 209–217.

11. W. Krull, *Ueber verallgemeinerte endliche Abelsche Gruppen*, Mathematische Zeitschrift, vol. 23 (1925), pp. 161–186.

12. ———, *Theorie und Anwendung der verallgemeinerten Abelschen Gruppen*, Sitzungsberichte der Heidelberg Akademie der Wissenschaften, 1926.

13. A. Loewy, *Ueber die Reduzibilität der Gruppen linearer homogener Substitutionen*, these Transactions, vol. 4 (1903), pp. 44–64.

14. ———, *Ueber die vollständig reduciblen Gruppen, die zu einer Gruppe linearer homogener Substitutionen gehören*, these Transactions, vol. 6 (1905), p. 504.

15. ———, *Ueber Matrizen und Differentialkomplexe*, I, II, III, Mathematische Annalen, vol. 78 (1917), pp. 1–51, 343–368.

16. C. C. MacDuffee, *On the independence of the first and second matrices of an algebra*, Bulletin of the American Mathematical Society, vol. 35 (1929), pp. 344–349.

17. F. D. Murnaghan, *The Theory of Group Representations*, Baltimore, 1928.

18. T. Nakayama, *Some studies on regular representations, induced representations and modular representations*, Annals of Mathematics, (2), vol. 39 (1938), pp. 361–369.

19. C. Nesbitt, *On the regular representations of algebras*, Annals of Mathematics, (2), vol. 39 (1938), pp. 634–658.

20. E. Noether, *Hyperkomplexe Grössen und Darstellungstheorie*, Mathematische Zeitschrift, vol. 30 (1939), pp. 641–692.

21. ———, *Nichtkommutative Algebra*, Mathematische Zeitschrift, vol. 37 (1933), pp. 514–541.

22. O. Ore, *Structures and group theory* II, Duke Mathematical Journal, vol. 4 (1938), pp. 247–269.

23. L. Pontrjagin, *Ueber den algebraischen Inhalt der topologischen Dualitätssätze*, Mathematische Annalen, vol. 105 (1931), pp. 165–205.

24. R. Remak, *Ueber minimale invariante Untergruppen in der Theorie der endlichen Gruppen*, Journal für die reine und angewandte Mathematik, vol. 162 (1930), pp. 1–16.

25. O. Schreier, *Ueber den Jordan-Hölderschen Satz*, Abhandlungen aus dem mathematischen Seminar der Hamburgischen Universität, vol. 6, pp. 300–302.

26. I. Schur, *Neue Begründung der Theorie der Gruppencharaktere*, Sitzungsberichte der Preussischen Akademie der Wissenschaften, 1905, pp. 406–432.

27. ———, *Beiträge zur Theorie der Gruppen linearer Substitutionen*, these Transactions, vol. 15 (1909), pp. 159–175.

28. B. L. van der Waerden, *Moderne Algebra*, 2 vols., Berlin, 1931.

29. ———, *Gruppen von linearen Transformationen*, Ergebnisse der Mathematik, vol. 4, 1935.

30. J. L. Wedderburn, *Lectures on Matrices*, American Mathematical Society Colloquium Publications, vol. 17, 1934.

31. H. Weyl, *The Classical Groups*, Princeton, 1939.

32. Zassenhaus, *Lehrbuch der Gruppentheorie*, Leipzig, 1937.

THE UNIVERSITY OF TORONTO,
TORONTO, CANADA.